ADVANCED DEEP LEARNING METHODS FOR BIOMEDICAL INFORMATION ANALYSIS (ADLMBIA), 2nd Edition

EDITED BY: Yu-Dong Zhang, Shuai Li, Carlo Cattani and Shuihua Wang
PUBLISHED IN: Frontiers in Public Health, Frontiers in Big Data and Frontiers in Artificial Intelligence

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-8325-4380-1
DOI 10.3389/978-2-8325-4380-1

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

ADVANCED DEEP LEARNING METHODS FOR BIOMEDICAL INFORMATION ANALYSIS (ADLMBIA), 2nd Edition

Topic Editors:
Yu-Dong Zhang, University of Leicester, United Kingdom
Shuai Li, Swansea University, United Kingdom
Carlo Cattani, University of Tuscia, Italy
Shuihua Wang, University of Leicester, United Kingdom

Publisher's note: This is a 2nd edition due to an article retraction.

Citation: Zhang, Y.-D., Li, S., Cattani, C., Wang, S., eds. (2024). Advanced Deep Learning Methods for Biomedical Information Analysis (ADLMBIA), 2nd Edition. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-4380-1

Table of Contents

frontiers in Big Data

EDITORIAL
published: 01 March 2022
doi: 10.3389/fdata.2022.863060

Editorial: Advanced Deep Learning Methods for Biomedical Information Analysis (ADLMBIA)

Yu-Dong Zhang[1]*, *Shuai Li*[2], *Carlo Cattani*[3] and *Shui-Hua Wang*[1]

[1] School of Computing and Mathematical Sciences, University of Leicester, Leicester, United Kingdom, [2] Department of Electronics and Electrical Engineering, Swansea University, Swansea, United Kingdom, [3] Engineering School (DEIM), University of Tuscia, Viterbo, Italy

Keywords: deep learning, biomedical information analysis, magnetic resonance imaging, social media, omics, surface electromyography, optical character recognition, computed tomography

Editorial on the Research Topic

Advanced Deep Learning Methods for Biomedical Information Analysis (ADLMBIA)

Due to numerous biomedical information sensing devices, such as Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Ultrasound, Single Photon Emission Computed Tomography (SPECT), and Positron Emission Tomography (PET), to Magnetic Particle Imaging, EE/MEG, Optical Microscopy and Tomography, Photoacoustic Tomography, Electron Tomography, and Atomic Force Microscopy, and so on, a large amount of biomedical information has been gathered in recent years. However, developing advanced imaging methods and computational models for efficient data processing, analysis, and modeling from the collected data remains a challenge.

Deep learning (DL) approaches have been rapidly developed in recent years, both in terms of methodologies and practical applications. DL techniques provide computational models of multiple processing layers to learn and represent data with multiple levels of abstraction. DL allows the capture of intricate structures of large-scale data implicitly and is ideally suited to some of the hardware architectures that are currently available.

The purpose of this Research Topic is to provide a diverse but complementary set of contributions to demonstrate new developments and applications of deep learning and computational machine learning to solve problems in biomedical engineering.

MRI and its related modalities are fundamental imaging tools in biomedical engineering. The method of Germuska et al. is based on the simultaneous acquisition of cerebral blood flow (CBF) and blood oxygen level-dependent (BOLD) weighted images during respiratory modulation of both oxygen and carbon dioxide. The authors present a machine learning implementation for the multi-parametric assessment of dual-calibrated fMRI data. Moreno López et al. evaluate two unsupervised approaches to denoise MRI in the complex image space using the raw information that k-space holds. The first method is based on Stein's Unbiased Risk Estimator, while the second approach is based on a blindspot network, limiting the network's receptive field.

Social media is now another critical tool that can provide information for biomedical analysis. Shah et al. use the advancement of natural language processing algorithms and large-scale data analysis. Their in-depth results show that the proposed method provides a viable solution in less time with the same accuracy compared to traditional methods.

Omics are novel, comprehensive approaches for analyzing the genetic or molecular profiles of humans and other organisms. Hassanzadeh and Wang use an integrated deep belief network to differentiate high-risk cancer patients from the low-risk ones in terms of overall survival. Their study analyzes RNA, miRNA, and methylation molecular data modalities from labeled and

OPEN ACCESS

Edited and reviewed by:
Thomas Hartung,
Johns Hopkins University,
United States

***Correspondence:**
Yu-Dong Zhang
yudongzhang@ieee.org

Specialty section:
This article was submitted to Medicine and Public Health, a section of the journal Frontiers in Big Data

Received: *26 January 2022*
Accepted: *28 January 2022*
Published: *01 March 2022*

Citation:
Zhang Y-D, Li S, Cattani C and Wang S-H (2022) Editorial: Advanced Deep Learning Methods for Biomedical Information Analysis (ADLMBIA). Front. Big Data 5:863060. doi: 10.3389/fdata.2022.863060

unlabeled samples to predict cancer survival and subsequently provide risk stratification.

The surface electromyography (sEMG) assesses muscle function by recording muscle activity above the muscle on the skin. In the paper of Shi et al., the knee flexion angle, hip flexion angle, ankle dorsiflexion angle, and sEMG signals of the seven muscles around the knee of three different data sets (walking data set, running data set, and walking and running mixed data set) are used as input of the 1D CNN. The attention mechanism is added to the network to observe the dimension feature that the network pays more attention to, thereby increasing the interpretability of the model.

Optical character recognition (OCR) technology is a solution for automating data extraction from printed or written text from a scanned document or image file and then converting the text into a machine-readable form to be used for data processing. The paper of Froese et al. builds a script that extracts real-time images from a medication pump and then processes them using Optical Character Recognition to create digital text from the image. This text is then transferred to an ICM + real-time monitoring software parallel with other retrieved physiological data.

A computed tomography (CT) scan combines a series of X-ray images taken from different angles around the body and uses computer processing to create cross-sectional images. The paper of Wang et al. builds a 12-layer convolutional neural network (12l-CNN) as the backbone network. Afterward, PatchShuffle is introduced to integrate with 12l-CNN as a regularization term of the loss function. Their model is named PSCNN. Moreover, multiple-way data augmentation and Grad-CAM are employed to avoid overfitting and locating lung lesions.

There is one review paper included in this Research Topic. Lan et al. introduce the origin, specific working principle, and development history of the generative adversarial network (GAN), various applications of GAN in digital image processing, Cycle-GAN, and its application in medical imaging analysis, as well as the latest applications of GAN in medical informatics and bioinformatics.

The ultimate goal of this Research Topic is to promote research and development of deep learning for multimodal biomedical images by publishing high-quality research articles, reviews, or perspectives, among other article types, in this rapidly growing interdisciplinary field.

AUTHOR CONTRIBUTIONS

Y-DZ was the guest editor of this Research Topic and wrote the editorial. SL, CC, and S-HW were guest editors of this Research Topic. All GEs approved the submission of this editorial.

ACKNOWLEDGMENTS

We thank the authors of the papers published in this Research Topic for their valuable contributions and the referees for their rigorous review.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Public Health

ORIGINAL RESEARCH
published: 14 January 2020
doi: 10.3389/fpubh.2019.00400

Assessing Canadians Health Activity and Nutritional Habits Through Social Media

Neel Shah[1], *Gautam Srivastava*[2,3]*, *David W. Savage*[4] and *Vijay Mago*[1]

[1] Department of Computer Science, Lakehead University, Thunder Bay, ON, Canada, [2] Department of Mathematics and Computer Science, Brandon University, Brandon, MB, Canada, [3] Research Center for Interneural Computing, China Medical University, Taiwan, China, [4] Northern Ontario School of Medicine, Thunder Bay, ON, Canada

When conducting data analysis in the twenty-first century, social media is crucial to the analysis due to the ability to provide information on a variety of topics such as health, food, feedback on products, and many others. Presently, users utilize social media to share their daily lifestyles. For example, travel locations, exercises, and food are common subjects of social media posts. By analyzing such information collected from users, health of the general population can be gauged. This analysis can become an integral part of federal efforts to study the health of a nation's people on a large scale. In this paper, we focus on such efforts from a Canadian lens. Public health is becoming a primary concern for many governments around the world. It is believed that it is necessary to analyze the current scenario within a given population before creating any new policies. Traditionally, governments use a variety of ways to gauge the flavor for any new policy including door to door surveys, a national level census, or hospital information to decide health policies. This information is limited and sometimes takes a long time to collect and analyze sufficiently enough to aid in decision making. In this paper, our approach is to solve such problems through the advancement of natural language processing algorithms and large scale data analysis. Our in-depth results show that the proposed method provides a viable solution in less time with the same accuracy when compared to traditional methods.

Keywords: data analysis, natural language processing, social media analysis, health analysis, machine learning, calories and physical activity

OPEN ACCESS

Edited by:
Yu-Dong Zhang,
University of Leicester,
United Kingdom

Reviewed by:
Shuai Liu,
Inner Mongolia University, China
Xiaochun Cheng,
Middlesex University, United Kingdom

***Correspondence:**
Gautam Srivastava
srivastavag@brandonu.ca

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *12 October 2019*
Accepted: *13 December 2019*
Published: *14 January 2020*

Citation:
Shah N, Srivastava G, Savage DW and Mago V (2020) Assessing Canadians Health Activity and Nutritional Habits Through Social Media. Front. Public Health 7:400. doi: 10.3389/fpubh.2019.00400

1. INTRODUCTION

Every year, Internet access is multiplying at a rate of 7% around the world (1). The level of yearly growth of social media users in Canada however is almost twice as high at 13%. Canada is a good representation for Internet usage with regards to the rest of the world as Canada had a 36.79 million people as of 2018, and among them, 33.05 million are Internet users. This is almost 90% of total population (1). As Internet access and quality increases, it creates an ideal condition for the growth of social media and other online activities. From 2017 to 2018 alone, Canadian's social media penetration reached 68% of the total population with 25.56 million people. The reason behind the exponential growth of social media users is mainly due to the technological advancement of smartphones and qualitative Internet services (with an average speed of Internet 45.64 Mbps in Canada) (1). This shows how deeply social media and the Internet has penetrated Canadian society.

An average Canadian spends approximately 6 h of their time every day on the Internet. Eighty-nine percent of the total population use the Internet daily for various activities (1). Smartphones are essential for social media, as they enable users to share their activities with ease of accessibility when compared to traditional social media devices such as computers. In current applications we see cameras integrated right into the application to upload content instantly without the hassle of conventional equipment. In Canada, smartphone users are growing with a rate of 6% every year which will increase the usage of social media, the Internet and different online services (1). This naturally generates an enormous amount of data and information which can be cultivated to form trends. Twitter has the most significant amount of activity among the social media platforms, at 7.2 million monthly active users all over Canada. The raw data collected includes all types of information from reviews of restaurants/products, political views, user's likes/dislikes, daily routines, just to name a few. Since Twitter provides a qualitative source of information that is a good measure of all social media platforms, in this paper our approach is to consider Twitter for analysis of public health as was originally shown in Dodds et al. (2). There are many factors which effect the quality of life and are complicated to measure directly. Presently, the best technique used to measure the quality of life are traditional surveys (3). However, the problem with these techniques include the types of data collected, actual collection of data, cost, degree of randomness, and time involved with the survey. These are all mitigating factors. Due to this conventional method, the chances of error are also increased. This in turn effects the decision of health policies and monitoring as it is not a proper representation but only a skim of the actual state of health within a given demographic area.

By studying the health of the population, trends can be formed with regards to prevalent health conditions. For example, diabetes, cancer, and heart conditions (4). Many of these health conditions are correlated with nutrition and level of daily physical activity. Health policy creators within government know this and conducts surveys, and programs to analyze the current health of nations (5). They can then use the information collected to put in appropriate policies and programs in order to help the population stay healthy and active.

The rest of the paper is organized as follows. In section 1, an introduction was given. Next, related works are discussed to get an idea of the similar current research being done in our field of health analysis in section 2. We then discuss the limitations in data analysis and how they can be solved in section 2.1. Our methodology is presented which includes: data cleaning, creating a database, phrase detection, and model training in section 3. Final results are given and form a detailed health analysis for Canada in section 4. Finally, this work concludes by summarizing our results and discuss the future possibilities in section 5.

2. RELATED WORK

Google Flu Trends was a real-time flu detection tool based on Google search query (6). If individuals search for a solution for the flu or any medical information related to the flu, the algorithm uses that information and considers their location as a potential flu affected area (6). However, the algorithm was proven to be ineffective. Paul and Drendze (7) gave a correlation when comparing cancer tweets, showing that there are higher obesity and tweets regarding smoking. They also found a negative relationship between health care coverage and tweets posted about diseases. With more sophisticated algorithms the accuracy of the data increases and this can be used to discover more true trends when looking at Twitter for health analysis.

Shawndra et al. (8) found that people who search about sodium content per recipe directly correlated with the number of people admitted in the emergency room of a major urban Washington hospital for congestive heart failure. Eichstaedt found that sentiment analysis of tweet language outperforms the traditional socioeconomic surveys for predicting heart disease at the country level (9). They correlated the growth of negative emotions in Twitter with the risk factor of heart disease on a large scale. This shows that social media analysis can be more effective than traditional surveys and may be the next step of methodology for future analysis done by the government.

Culotta et al. (10) analyzed tweets which contain the daily habits of the account holders. The results were a "deep representation" of the US community in regards to their daily negative engagement concerning their routine such as watching television, playing, or reading. Abbar also did an analysis of data on Twitter for caloric analysis at the country level. They classified food-related tweets and found the caloric value of such food. This analysis gave a brief understanding of the food habits of the people in different demographic areas (11). Subsequently, Lexicocaloricmeter (`LCM`) became one of the most sophisticated approaches toward the health analysis of people at the country level. This is done by utilizing social media. `LCM` is an online instrument that is designed for measuring social, physical, and psychological examination at a large scale. Sharon et al. (12) developed it for public health monitoring and to create health policies through data-centric comparison of communities at all scales. Oversimplification exists in data analysis which means that the data is being classified in basic categories. Doing this results in looking only at the data present instead of looking deeper into the meaning or relevance of the data. This methodology is known to cause bias. An example of this is a piece of data from a Twitter account that says "**the test was a piece of cake**." This idiomatic expression that has very little to do with food. Instruments like `LCM` will take this data as a food tweet and add it to is trends. This causes errors and inaccurate trends which needs to be addressed in future models. Models need to have a resistance to oversimplification. `LCM` extracts text related to caloric input and caloric output and calculates their caloric content (13, 14). They also use food phrases from a 450-plus database and physical activity phrases from a 550-plus database. The second step is to group categorically similar words and phrases into small pieces called lemmas. They then assign caloric values to it, based on the food and physical activity. To get these lemmas, they use a greedy selection algorithm. Food caloric value is represented as C_{in} and activity caloric value is represented as C_{out}. C_{rat} is calculated as

shown in Equation (1).

$$C_{rat} = \frac{C_{in}}{C_{out}} \quad (1)$$

To find the average caloric value of different provinces or countries, frequency of all food and activity related words is counted and then caloric values to all words are assigned. Next, the standard C_{rat} formula is used to compute the caloric ratio of each place. The authors consider 80.7 kilograms as the average weight for metabolism equivalent of tasks; this is subtracted from the calorie's physical activity value.

2.1. Limitations

For simplicity, the LCM did not use any filter for tweets beyond their geographic locations. This causes bias in the data-set because the user may live/eat in different locations. This also causes the users eating habits to affect another location's data-set instead of affecting there home location data-set. For example a given user might be from Toronto and go on a trip and eat in Montreal. With LCM's current filter the user's data will affect the data-set gathered from two separate locations instead just Toronto as it should. This causes a loophole in the data-set that will cause inaccuracy.

LCM's data-set is quite limited with only 451 food phrases (15). The food phrase data-set has the most common food names which limits its applicability. Also when people talk about the food, it can be called anything such as the name of special dish in a certain restaurant (15). Different cultures have different foods and this is very important in a country as diverse as Canada. So the database of food phrases in our model must be large in order to accommodate for all possibilities in order to be accurate.

Another limitation of LCM is that the Twitter account may talk about food or an activity in a metaphorical perspective. Food words are commonly used in idiomatic expressions in the English language. Some examples include: "**bring home the bacon**," "**crying over spilled milk**," and "cup of tea." LCM will still consider these phrases as food items in their system and assign values to them. The approach used in LCM cannot solve this problem, and therefore creates bias in the system. An example of this includes if a person tweets the phrase "**you are the apple of my eye**," the present algorithm will consider apple as a food. But in this case, it is not related to food. Also, a lack of Natural Language Processing (NLP) understanding of such approach creates higher chances for the bias output (16). Due to this, unnecessary data will enter the data-set and create false trends, over-fitting and decrease accuracy of the overall analysis.

3. METHODOLOGY

In the design of our system, the focus of the system was placed on large scale analysis of social media data with regards to health analysis. The focus of the system is also on training a NLP model based on a large amount of data that is processed to get factual information about the health of Canadians. **Figure 1** shows the architecture of our health analysis system. It is divided into two subsections, namely training (offline mode) and the analysis component (active system). As shown in **Figure 1**, the first step is to collect all the raw data. To manage and process data, Elasticsearch system is used which is designed and developed through Elasticsearch locally at Lakehead University's High-performance computing facility (17). It can handle and analyze Terabytes of text data with a low time overhead. This helped collect the necessary data very efficiently from the pool of data. Once the pool of data was collected, the next step was data cleaning.

3.1. Data Cleaning

BOX 1 | Data cleaning examples.

Step 1: Emoji or Emoticons gave there respective meaning
Original text: "I am getting 2 old to be mango Gonna retire soon and be joesh #ROFL :-)"
Processed text: "i am getting 2 old to be mango gonna retire soon and be joesh #ROFL Happy face smiley"

Step 2: Covert all text in lower-case characters
Original text: "I am getting 2 old to be mango Gonna retire soon and be joesh #ROFL"
Processed text: "i am getting 2 old to be mango gonna retire soon and be joesh #rofl happy face smiley"

Step 3: Removing stop words
Original text: "i am getting 2 old to be mango gonna retire soon and be joesh #rofl"
Processed text: "getting 2 old mango gon na retire soon joesh #rofl happy face smiley"

Step 4: Removing special characters
Original text: "getting 2 old mango gon na retire soon joesh #rofl"
Processed text: "getting 2 old mango gon na retire soon joesh rofl happy face smiley"

Step 5: Removing numbers
Original text: "getting 2 old mango gon na retire soon joesh rofl"
Processed text: "getting old mango gon na retire soon joesh rofl happy face smiley"

Data cleaning is crucial when dealing with user's raw data such as tweets, feeds, or chats, as shown in **Box 1**. Raw data is not structured or cleaned unlike typical formats such as blogs or essays. When tweets are written, it may include hashtags, slang words, emojis, emoticons, and unstructured data. Because of that, raw data is used as a feature in the model as an input; to make this data more sensible and more reliable for the model it must be cleaned.

In the first step of data cleaning, all Emojis or Emoticons are converted to there respective meaning through the "emot" open source library given in Shah (18). It helps to understand the text when a name that is not in the database. When emojis are related to food, it will be easy to understand that the tweet is related to food. In the next step, all text is converted into lower case which makes word matching and processing easy in further processes. Step 3 is removing stop words which help to eliminate unnecessary features in our model (*the*, *a*, *an*, *when*, *what*). The

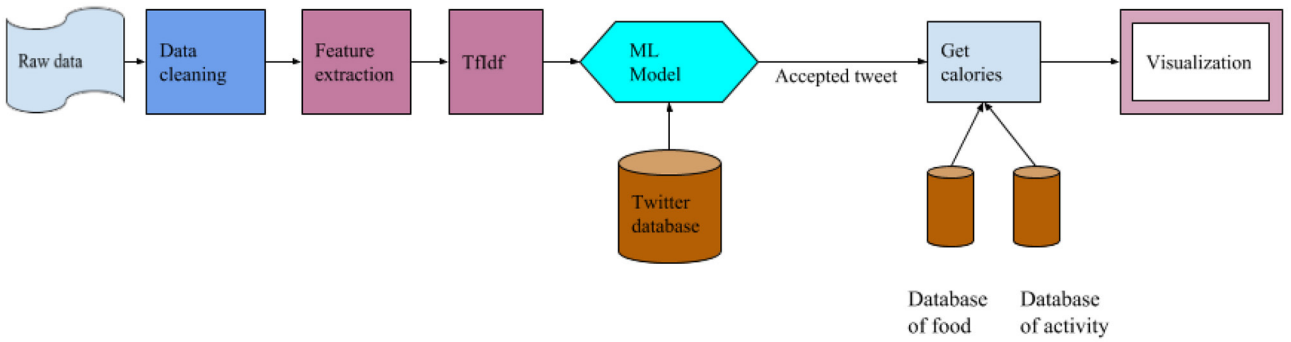

FIGURE 1 | Architect of the analysis system.

next step is to remove special characters. Finally, numbers are removed. This is because numbers are not useful for identifying whether the text is related to food or not and it is also not a source of information for our analysis. Removing all unnecessary text or data will limit the size of the feature matrix and speed up the training and classification task as some features are directly propositional to the speed of model training. As the number of features increases the speed of training the model also increases as given in Batista et al. (19). Here, hashtags are not removed because it gives valuable information. For example, when users talk about specific foods which are not common but use hashtags alongside text, for example #burger#delicious, then the user is talking about a burger or some other food which can be quickly identified. So, special characters were removed while keeping hashtagged text.

3.2. Database

To calculate the caloric value, two types of data-sets are needed: first for food and its caloric values and the second for activity and its caloric-burn value. When a data-set is gathered for food, our research found out that there is not even a single data-set available which includes the different types of food items and their nutrition values. At present, the Canadian Food Nutrient Database and USDA Food Composition Databases are the main sources of information related to food and nutrition facts of the food in Canada. But the limitation with these databases is the lack of data in terms of specific foods/items such as "**Chicken masala**" or "**Penne arrabiata**." Usually, people tweet about specific items they eat during their meal at a restaurant or any other place. It means the present data-set is very domain-oriented for things such as fast food, vegetables, or frozen foods, but they will not contain all the major types of food that people talk about on social media as shown before. There are other problems after getting the data-set. First, to find what type of specific food the users talk about. Next, to find the caloric value of that specific food. To solve these problems, a new data-set is needed that combines different food domains which contain all major foods and their different nutrition values. This is why, "**Food in one**" dataset was created which includes a combination of all open source data-sets such as the Open Food Facts which is a major source of food

TABLE 1 | Food database.

Name	Data
food_name	Name of the food
food_ingredients	Ingredients use to make the food
fat_100g	Fat per 100 g of food
energy_100g	Energy value per 100 g of the food
carbohydrate_100g	carbohydrate value of that food at 100 g

names, Canadian Nutrient File, and USDA Food Composition Databases. **Table 1** shows the structure of current food data-set.

Our newly created data-set contains 338, 889 foods with all the required information. This is an open source database available at DataLab. This includes all different types of major food sources like fruits, vegetables, fast food, and regular food. In our data-set, more than 70% of the food items are from the USA, Canada, and France. This is because our focus is mainly on Canada's health situations and these are the main sources of food in the Canadian market.

To understand the nutritional value of all food items in the database a Normalized Kernel Density Estimation KDE is used. **Figure 2** is the KDE diagram of all the food that is present in the data-set. This is along with their nutrition values including fat, carbohydrate, and energy per 100 grams. As can be seen in the second bar chart energy values mostly lie between mid-range while the fat bar chart has diverse values from an extreme high to an extreme low. This represents the diverse nature of our data-set that includes various type of foods.

Figure 3 shows the normalized KDE graph of nutrition values of vegan and non-vegan food, where orange represents the vegan food, and blue represents the non-vegan food. The results show that the distribution is quite similar for products with "Vegan" labels. As shown in the **Figure 3** non-vegan food has high fat and energy values when compared to vegan foods on average. While the scatter graph, between carbohydrates and fat, shows a vegan diet has a lower energy value when compared to non-vegan foods with regards to the same amount of carbohydrate. The last raw scatter graph shows that all our food is categorized as vegan or non-vegan food.

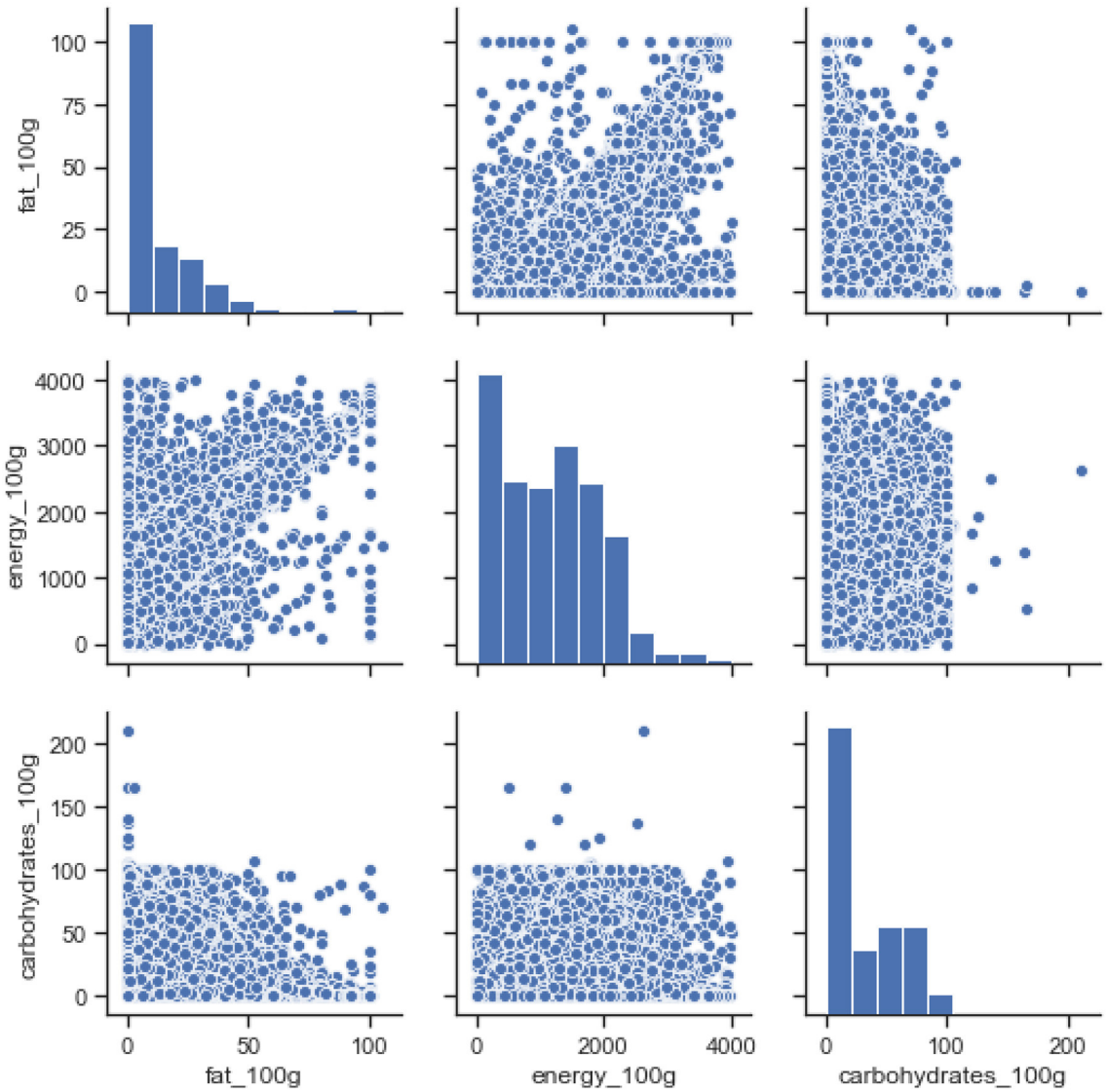

FIGURE 2 | Nutrition value of all food types in the database.

To differentiate between vegan and non-vegan food, our methodology involved first finding out if the word "vegan" is present beside the name of the food in the database as shown in Algorithm 1. Then vegetables, fruits, and juice were added as well to the vegan food category. Any other foods are considered as non-vegan food.

Figure 4 shows the KDE graph of carbohydrates per 100 g concerning the distribution between vegan and non-vegan foods. It also shows that some non-vegan food has high carbohydrate content than vegan foods. While, in other aspects of nutrition, the gap between vegan and non-vegan food is not so big.

Figure 5 shows the KDE graph of fat per 100 g distribution between vegan and non-vegan foods. It also shows that some

Algorithm 1: Identifying food is vegan or non-vegan

Input: List of food name in data-set
Result: Food is Vegan or Non-Vegan

```
FoodDataset;
while Food_in_FoodList do
    name = Food;
    if name contains "vegan" then
        flag = True;
    else
        flag = False;
    end
end
```

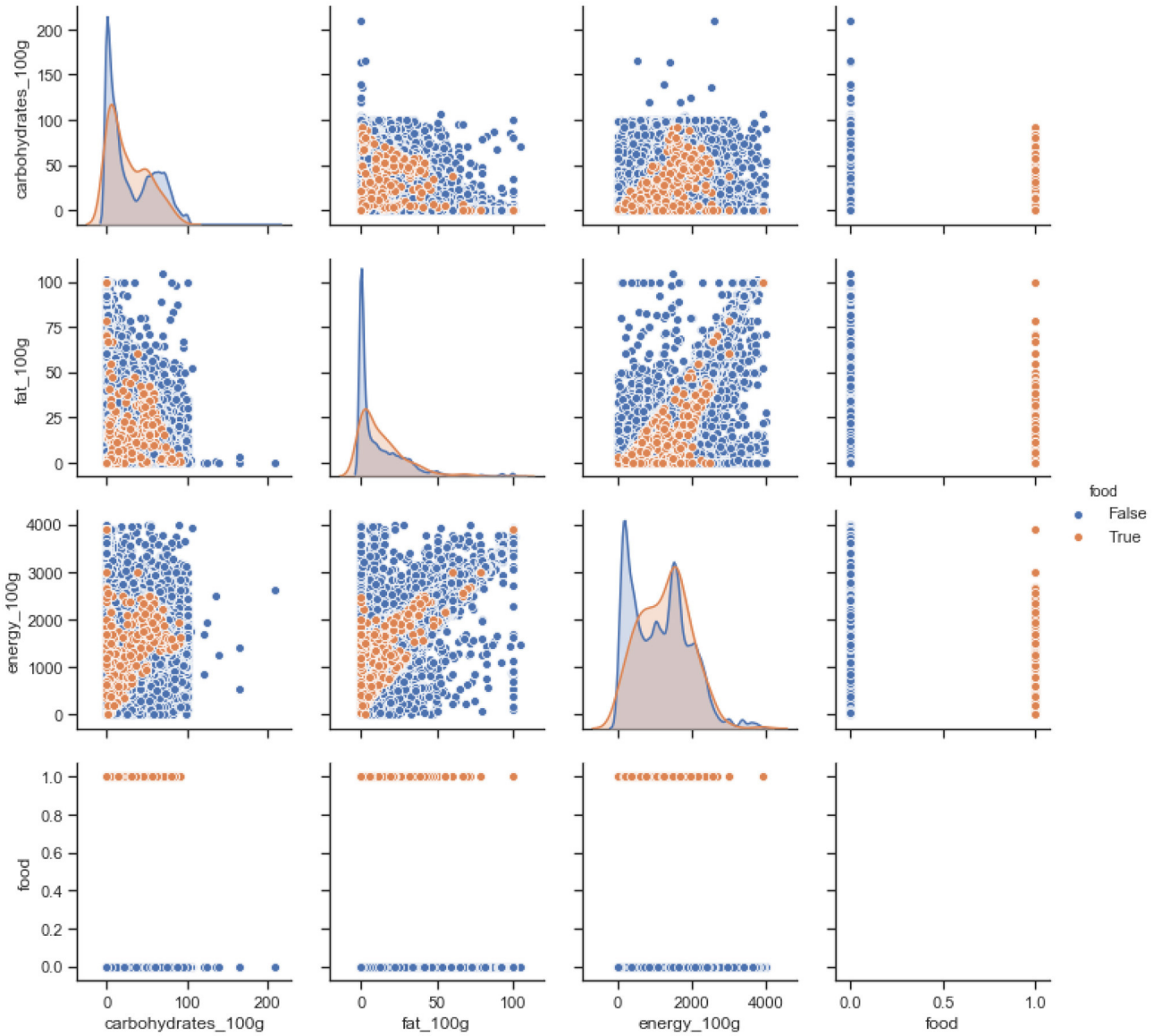

FIGURE 3 | Nutrition values of vegan vs. non-vegan food in the database.

non-vegan foods have sharply high-fat content than vegan products. In other aspects, there is not much of a big difference between vegan and non-vegan foods, for example, fat content.

Figure 6 shows the KDE graph of energy per 100 g distribution between vegan and non-vegan food.

In order to analyze public health, the second database needed is an activity database, where the average activity time is taken and related to its caloric values. The most common activities that are posted by people on social media were chosen. The database now contains 1,400 different activities and their caloric values that are available. To calculate the average caloric value, fixed weight, and metabolism were used with average Canadian weight at 80.3 kg. **Table 2** shows the attributes of the activity data-set.

To analyze public health on a large scale, the Twitter data-set was considered as the primary source for data. This will be used to do basic querying and analysis of the system at a large scale; Elasticsearch based analysis system is developed for real-time querying and searching of Twitter data (17). From that system, 99,999,986 tweets were analyzed between 2018 and 2019.

3.3. Phrase Detection

In social media, a text data phrase gives more information than a single word. In previous work, one of the limitations is the inability to understand the phrases of multi-words. For example, when anyone tweets "**you are the apple of my eye**," it considers "**apple**" as a food item. In our system

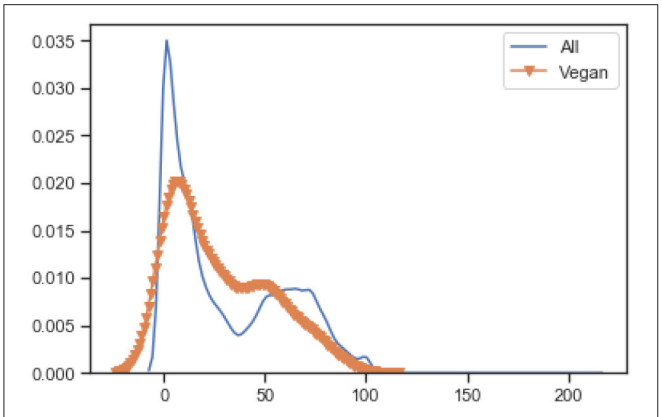

FIGURE 4 | KDE of carbohydrates per 100 g.

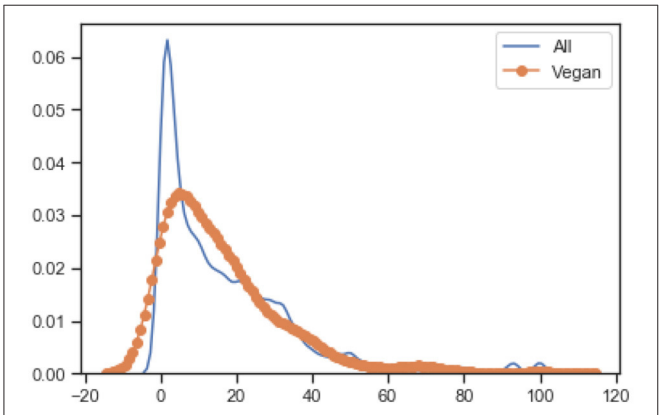

FIGURE 5 | KDE of fat per 100 g.

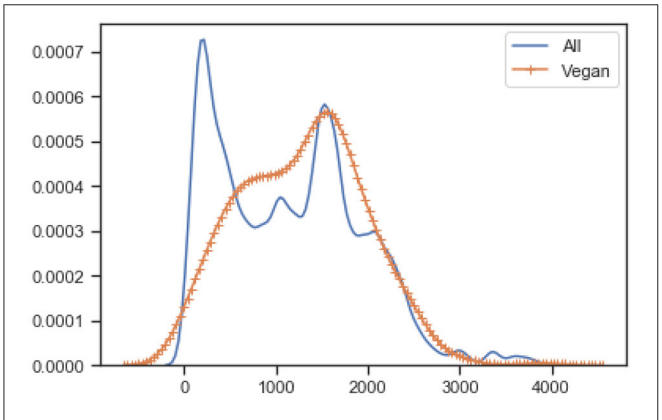

FIGURE 6 | KDE of energy per 100 g.

apple of my eye is considered as a single phrase, which has a specific meaning. One common example as well is "**you are a smart cookie**" where "**smart cookie**" has meaning as a phrase. To overcome this limitation new features are added during the training of the NLP algorithm. This algorithm was developed initially by Williams (20), and it is based on the distance between two words. This text partitioning algorithm is based on William's fine-grained text segmentation algorithm. It considers the whole text as two parts: word and non-word tokens. The important feature of this algorithm is that it considers non-word tokens as a linker between two words. For example, in the phrase "**apple of my eye**," "**of**," and "**my**" are non-words and works as a joiner of the single phrase with singular meaning.

Algorithm 2 is a phrase detection algorithm that uses a concatenation operation. This links the tokens together to create forms and then finds out how related the form is to the lexicon. If the form is not correlated with the lexicon then the next possible form is analyzed. If the form is related to the lexicon, it is considered as a phrase.

Algorithm 2: Phrase detection algorithm

```
Input: List of words of the text - tokens
Result: List of tokens as Phrase - lexemes
phrase_detection(tokens):
lexemes[]
N = length(tokens)
while N do
    index = (N+1):1
    foreach i in index do
        form = join(token[0:i])
        remaining = tokens[i:N]
        if form related lex then
            lesemes = lexemes.add(form)
            if length(tokens)=1 then
                pass
            else
                tokens = remaining
            end
        end
        break
    end
end
return lexemes
```

Algorithm 2 is based on Boundary-based multi-word expression segmentation with text partitioning by Williams (20). This algorithm focuses on the next possible word pair, which means a lower precision and efficiency for complex bound phrases. But the phrase information will be derived from a gold standard data-set. For example, Supersense-tagged Repository of English with Unified Semantic and Riter and Lowlands data-set of superscience-annotated tweets for the SemEval in Williams (20). Due to that pre-information of the phrases finding, results with simple and common phrases are easy. **Box 2** shows the results of the phrase extraction algorithm used.

BOX 2 | Phase extraction results.

text: "I saw the sweet potatoes."
phrase: "['sweet', 'potatoes']"

text: "My daughter is an apple of my eyes."
phrase: "['apple', 'eyes', 'daughter']"

Our results shows that the phrase detection algorithm will analyze the text and predict phrases for example the phrases "**sweet potatoes**" and "**apple eyes daughter**" are a single phrase. Even though our data cleaning process removes stop words the phrase detection algorithm can still detect a complicated sentence like "**my daughter is the apple of my eye**" as the phrase "**apple eyes daughter**."

3.4. Model Training

In the next step, the machine learning model was trained for classification of the tweet. As shown in **Figure 7**, our first step is to take the raw data and clean it (explained in section 3). After cleaning the data, feature engineering was performed. In feature engineering, term frequency-inverse document frequency (`tf-idf`) is created and phrase extraction occurs. Here, phrases as a single word are used as a feature in `tf-idf`. After that, the model was trained and then used as a pre-trained model for binary classification of the tweet for food and non-food related tweets.

Two types of features were used: `tf-idf` with phrases and word embedding with 2 g. `Tf-idf` is used for Naïve Bayes (`NB`), Logistic Regression (`LR`), and Random Forest (`RF`), and Support Vector Machine (`SVM`). Embedding for Shallow neural network (`SNN`), Convolutional neural network (`CNN`), and Reinforcement neural network (`RNN-GRU`) was used.

TABLE 2 | Activity database.

Name	Data
activity_name	Activity name
caloric_value	Caloric value of the activity

4. ANALYSIS AND RESULTS

In previous works, researchers tried to find the caloric value through non-`NLP` or basic `NLP` algorithms. Because of that, the false positive rate of data was high, and this decreased the accuracy of the results. False positive errors will start to increase as the data amounts increase (21). This affects the accuracy and accountability of the system. Many models are available for the classification of text. In this paper the binary form of classification is much easier than multi-class classification. It also removes the necessity of the necessity to use advanced deep learning algorithms. The first algorithm tested was `LR`. This measures the relationship between one or more categorical dependent and independent variables. It will be estimated through logistic (sigmoid is more common presently) function.

Figure 8 shows the confusion matrix of `LR`. It also shows that 92% of food tweets were successfully identified. While recognizing only 85% of the non-food tweet as non-food. But the false positive ratio is very high, 15% which landed into more noisy data.

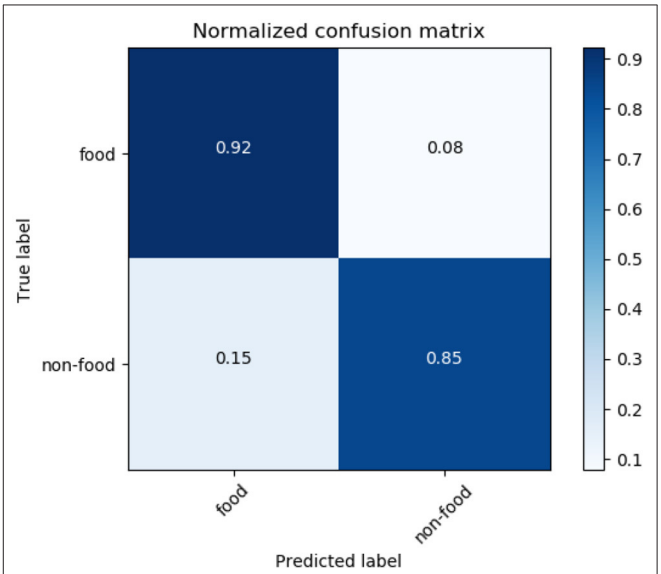

FIGURE 8 | Logistic regression confusion matrix.

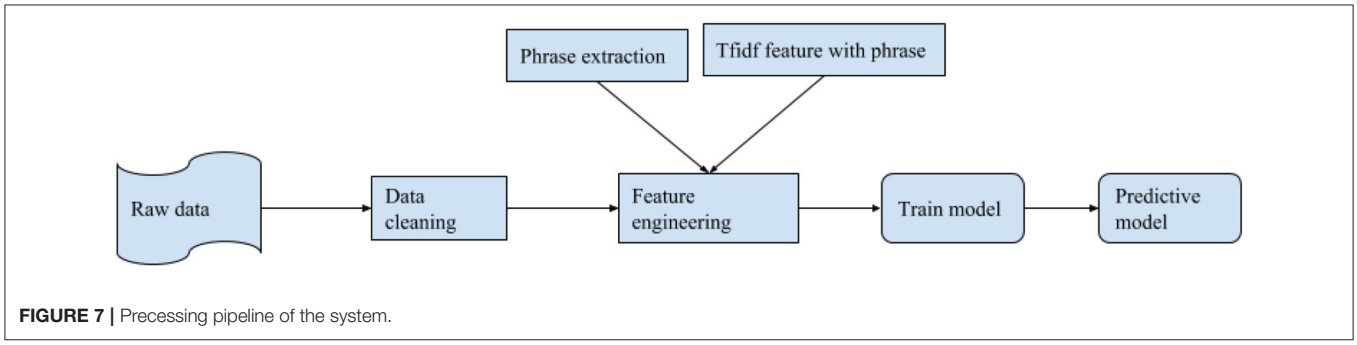

FIGURE 7 | Precessing pipeline of the system.

Figure 9 shows the training curve of LR. As shown, as the number of training samples increases its accuracy is also increasing.

Our second algorithm is the NB algorithm with `tf-idf` features on a word level. This classification of algorithm techniques is based on Bayes' theorem. This assumes Independence between predictors. Meaning, it implies that one feature in the model is unrelated to another feature.

Figure 10 shows the confusion matrix of NB. It also shows that the methodology can successfully identify 91% of food tweets as food tweet, while only 61% of the non-food tweets as non-food tweets and a high rate of false positive as 39%, which is quite high to get accurate results. **Figure 11** shows the training curve of NB. It shows the accuracy of the algorithm increases as the sample size is increased. After 400 K samples, the accuracy of the algorithm is almost at 80%.

The third model analyzed was the RF model. This is a type of bagging model, and it is a part of the tree based model. An advantage of this model is that it gives more accurate pr editions when comparing it to any simple CART or regression model in specific scenarios. **Figure 12** shows the confusion metric of RF. It also shows that our methodology successfully identified 97% of the food tweets as food tweets, while 88% of the non-food tweets are recognized as non-food tweets. On the other hand, the false positive rate is also as low at 12%. This result shows the highest accuracy among the other tested models. **Figure 13**

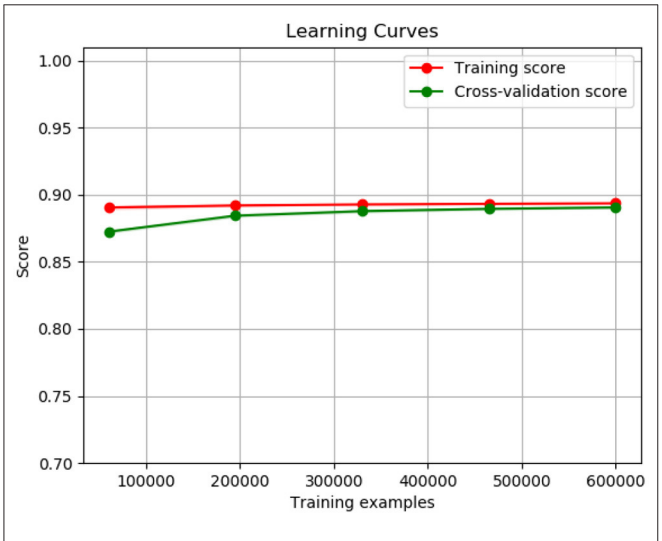

FIGURE 9 | Logistic regression training curve.

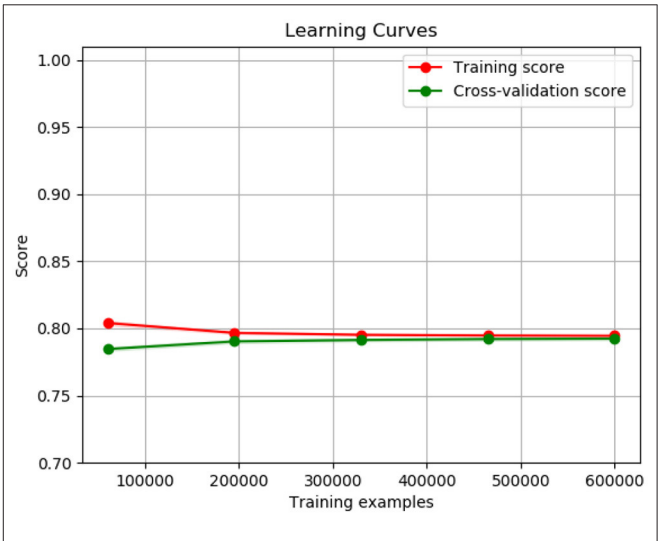

FIGURE 11 | Naïve Bayes training curve.

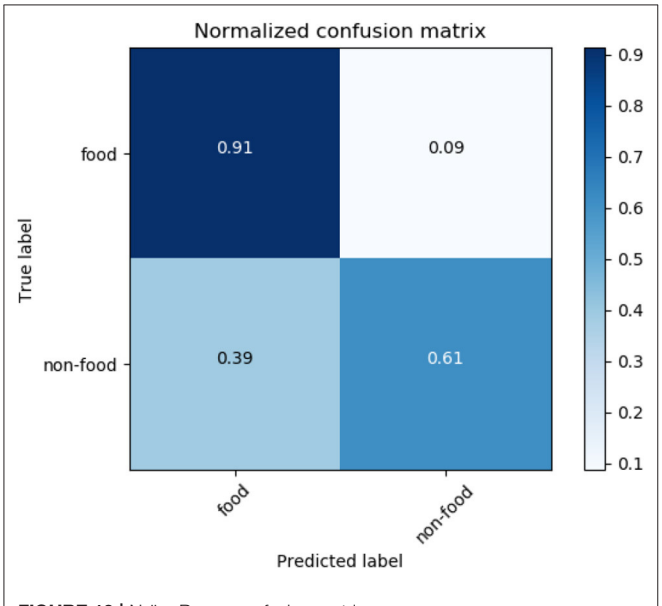

FIGURE 10 | Naïve Bayes confusion matrix.

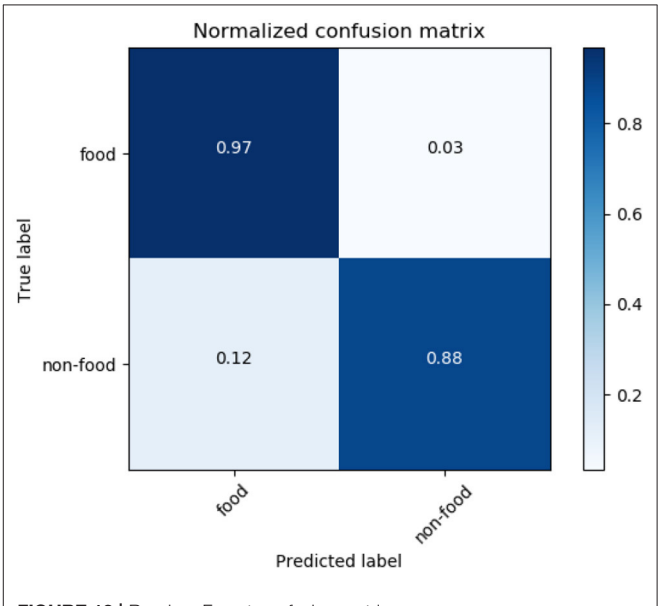

FIGURE 12 | Random Forest confusion matrix.

shows the learning curve of RF. It also shows that the accuracy of the algorithm increases as the data size increases.

Table 3 shows an overview of the accuracy of the different models used to test our methodology. From the above analysis, the RF model gives us the best results for the binary classification of tweets in food and non-food categories. The next step is to get information about the calories and the user's activity based on our data-set. Three different values became our focus: caloric value they gain from their food (C_{in}), the caloric value they burn from their activity (C_{out}), and their caloric ratio from the first two values (C_{rat}) (Equation 1). Those three values are co-related to the 37 measures of the well being and health. In LCM, the authors found a statistically strong correlation between high blood pressure, inactivity, diabetics, and obesity rates (12). Now to count these three values C_{in}, C_{rat}, C_{out} our methodology depends on the benefits from each individual tweet.

Algorithm 3 gives the value of C_{in} value for C_{rat} as shown in Equation (1). In the first step, tweets in our data-set are taken to produce the caloric value of each tweet using the Calorie_consumption function. In that function, text is used to locate any food phrases. In the second step, the individual words or phrases that are found in the tweet are compared to the food data-set. If present, caloric value is taken of the word or phrase from the food dictionary. Then calories of the given food are multiplied with the frequency of that word in the text. It is then stored in the Frequency matrix and to get a normalized C_{in} value. The sum of all calorie values from all the tweets are divided by the sum of the frequency matrix.

Algorithm 4 gives the value of C_{out} value for C_{rat} as shown in Equation (1). In the first step, tweets in our data-set are taken to get the caloric values of each possible word in the tweet. Next, each word from each tweet is checked with the activity data-set. The data-set will give a caloric burn value for each tweet. To normalize the C_{out} value, the summation of caloric values is divided by the frequency of each activity phrase. To count the C_{out} values, how many calories a person can burn is stored from a particular activity. For that, a body weight of 80.7 kg is assumed as the standard average weight of a Canadian adult.

Table 4 is the result of the 100 K tweets gathered between 2018 and 2019. 50 K tweets were chosen pertaining to food and 50 K tweets pertaining to activity from each province and territory randomly, which combine to form our 100 K tweet data-set (22). The Twitter API was used to collect the data without any filters and therefore makes our collection of tweets random (23). **Table 4** shows the top 10 foods in Canada, and it clearly shows that junk food and hot drinks are the most common.

Algorithm 3: C_{in} algorithm

Input: Tweet text (T), list of food, dictionary of caloric value of food with food name as key
Result: C_{in}

```
Global Frequency_Matrix = {}
Global FoodList = list_of_food
Global FoodDict = dictionary of caloric value of food with food name as key

Calorie_consumption(text):
phrase = phrase_detection(text)
while word in text do
    if phrase in FoodList then
        cal = FoodDict[phrase]*Frequency(phrase);
        Frequency_Matrix[phrase]++;
    else
        if word is FoodList then
            cal = FoodDict[word]*Frequency(word);
            Frequency_Matrix[word]++;
        else
            return 0
        end
    end
    return 0
end

calorie = 0
Get_Cin(Tweets):
while tweet in Tweets do
    calorie = calorie + Calorie_consumption(tweet)
end
return Cin= calorie/sum(Frequency_Matrix)
```

Algorithm 4: C_{out} algorithm

Input: Tweet text (T)
Result: C_{out}

```
Global Frequency_Matrix = {}
Global ActivityList = list_of_activity
Global ActivityDict = dictionary of caloric value of activity with activity name as key.

Calorie_burn(text):
while word in text do
    if word in ActivityList then
        cal = ActivityDict[word]*Frequency(word);
        Frequency_Matrix[word]++;
    else
        return 0;
    end
    return 0
end

calorie = 0
Get_Cout(Tweets):
while tweet in Tweets do
    calorie = calorie + Calorie_burn(tweet)
end
return Cout= calorie/sum(Frequency_Matrix)
```

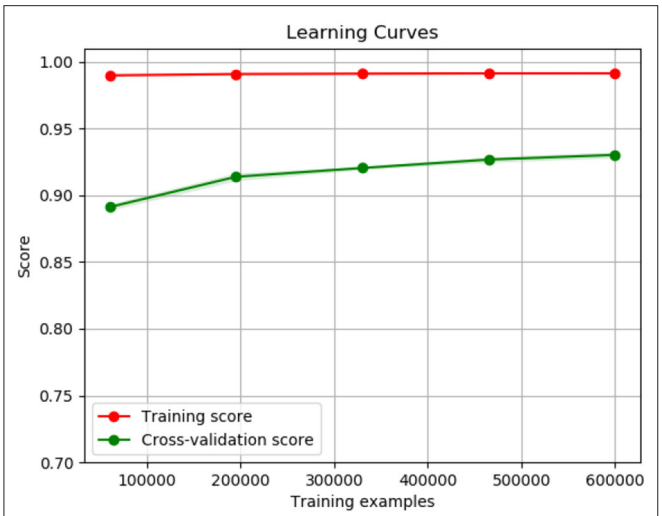

FIGURE 13 | Random Forest training curve.

TABLE 3 | Model accuracy.

Model	Accuracy %
Naïve Bayes	79.202
Linear regression	89.155
Random Forest	93.406
CNN	60.142
RNN-GRU	60.034
SVM	56.031

TABLE 4 | Top 10 food in Canada.

Rank	Food	Number of tweets
1	Coffee	38,785
2	Burger	35,166
3	Pizza	34,369
4	Noodles	27,891
5	Cake	18,456
6	Pie	17,982
7	Juice	16,711
8	Tea	16,631
9	Fruits	15,987
10	Veggies	11,473

Table 5 is the list of the top 10 activities in Canadian tweets. It clearly shows that large number of people are choosing to watch something regularly instead of physical exercise. It also shows that walking and running are the most common exercises people do. This means physical inactivity is increasing throughout Canada.

Figure 14 shows the most common foods people tweeted about in different provinces and territories of Canada. Clearly in Ontario and Alberta, the most common foods tweeted are Pizza, while Quebec's most common food is Fries. It also shows that coffee is the most common tweeted drink in provinces like Manitoba, Saskatchewan, Yukon, and Northwest Territories as compared to tea, which is more common in tweets originating from British Columbia.

TABLE 5 | Top 10 activity in Canada.

Rank	Activity	Number of tweets
1	Watching (seeing)	42,489
2	Reading	31,762
3	Walking	28,127
4	Running	27,838
5	Drinking	27,339
6	Sitting	24,347
7	Cooking	22,561
8	Skiing	18,947
9	Gym	16,585
10	Playing	14,191

Figure 15 is about the most common activities people tweeted about in different provinces and territories of Canada. The result shows that watching (TV) is the most common activity in dense populated provinces of Canada. This includes Ontario, Quebec, Alberta, Yukon, and Northwest Territories. This shows that there is less physical activity among people in these provinces, which is an alarming situation when looking at the individual's food consumption versus activity they do to burn calories.

From the Canadian government's own health analysis (24), it shows that Ontario and Quebec have 38.3 and 23.2% of the total population respectively. When both are combined, it is 61.5% of the total population. Caloric ratio is highly correlated to blood pressure and obesity, which according to LCM that 77.92% population has a higher chance of getting obese and/or higher blood pressure which is a staggering percentage. Our results here correlate directly to the results given by the Canadian Institute of Health Information report "**Obesity in Canada**" (25). It also shows the rapid growth of Obesity in Ontario and Quebec.

Figure 16 shows the Caloric ratio based on Equation (1). If the ratio is >1 that means that the province's consumption is greater than the caloric usage. The opposite is true when the ratio is <1. According to **Figure 16**, Yukon, Newfoundland and Labrador, and Saskatchewan's caloric consumption is higher than their caloric burning at this instance. Northwest Territories, Manitoba, and British Columbia have a caloric burn that is higher than their caloric consumption. Caloric ratio is highly correlated to blood pressure and obesity (12). When looking at **Figure 16**, 77.92% population has a caloric ration >1.0. This can cause greater chance of getting obese, and higher blood pressure. This population estimation is based on the population numbers from the 2016 Canadian Census and was calculated by adding up the populations from each individual province with a 1.0 caloric ratio or higher. This is alarming because it represents such a huge part of Canada's population.

Obesity percentages for the provinces of Canada are analyzed in **Figure 16**. When looking at the Obesity in Canada report, published in 2017, the order of provinces from lowest obesity rate to highest obesity rate is the following:

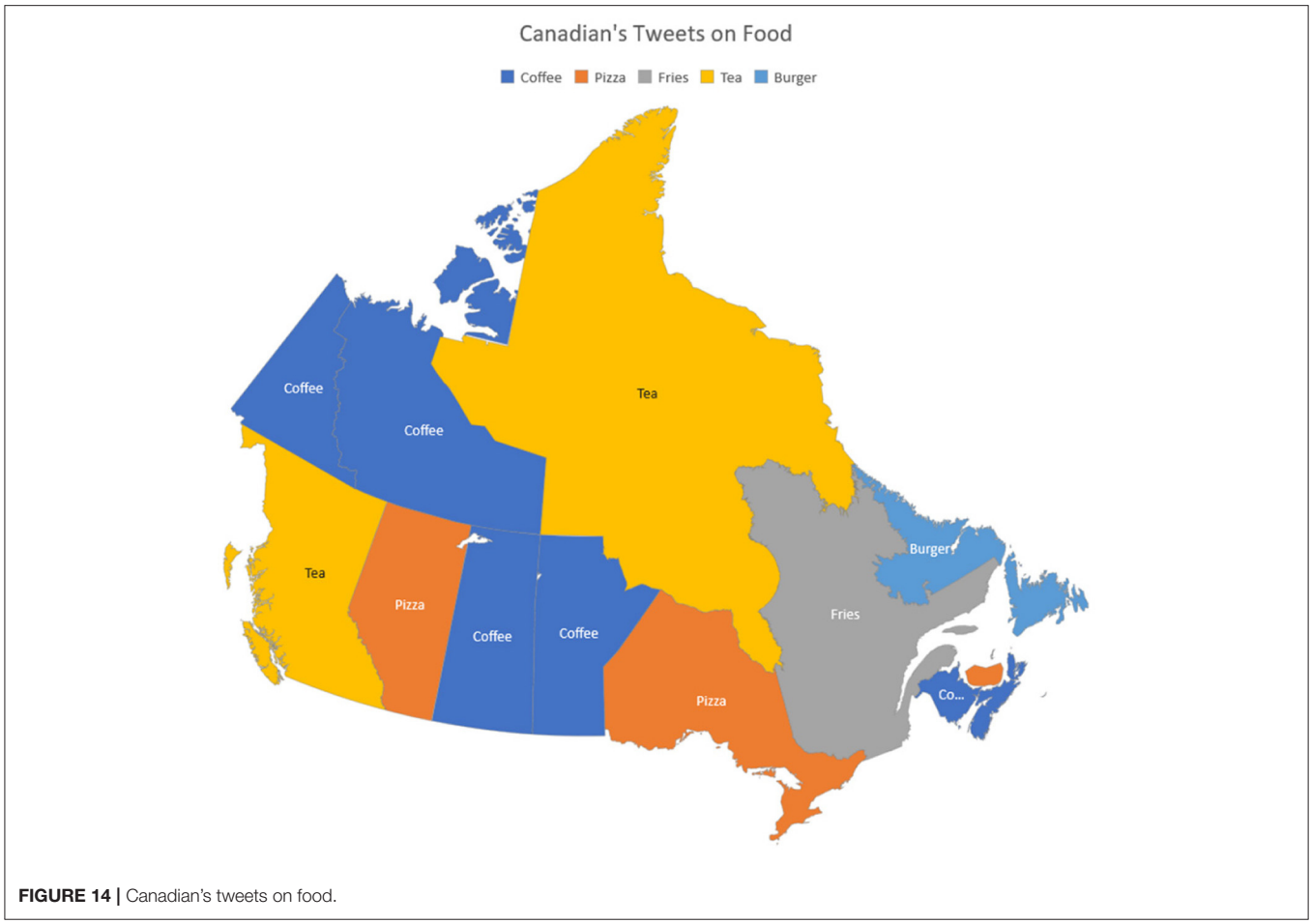

FIGURE 14 | Canadian's tweets on food.

1. British Columbia
2. Quebec
3. Ontario
4. Alberta
5. Manitoba
6. Saskatchewan
7. Newfoundland and Labrador (25).

When comparing that with the lowest to highest rank in **Figure 16**, British Columbia is clearly the lowest, while Quebec and Ontario are tied, and Alberta and Manitoba are switched in ranks, Saskatchewan and Newfoundland and Labrador are the highest. The significance of the comparison above is that it shows a strong correlation between **Figure 16** and the data from the report. The provinces with the highest ratios also have the highest obesity rates and the opposite is also true. This goes as far as to show that our data can be just as reliable as a published Canadian report. By using this knowledge, the Canadian government can target healthy living programs in the provinces that need it like Newfoundland and Labrador.

Figure 16 only shows one caloric ration per province from the tweets collected at that instance. In order to become obese you need to constantly have a ratio >1 over the span of weeks. That way you are consuming more calories that you are burning which leads to the gaining of weight. Therefore, our **Figure 16** can be considered as one data point in an obesity rate trend and in the future it can be added with many other points taken in different times to accurately show the rate of obesity in Canada. By this logic you can also see the trend from one point to another and immediately see if the programs implemented to counteract obesity have worked. The result is going from tweets to a real time analysis of the rate of obesity in Canada.

5. CONCLUSION AND FUTURE WORK

Developing high-performance machine learning model with a limited amount of training data is always a challenge, as it restricts the use of more complex deep learning and neural models. Our model gives 93.406% accuracy in binary classification of food and non-food tweet. This result shows that social media analysis on a large scale with the use of better NLP algorithms can help us to identify food and activity related tweets more accurately. This helps us to gain a larger perspective on daily activities and its effect on people's health. Our results convey a complex relationship between health and social media. The presented approach is faster when compared to traditional survey methods causing data to be readily available as well a close

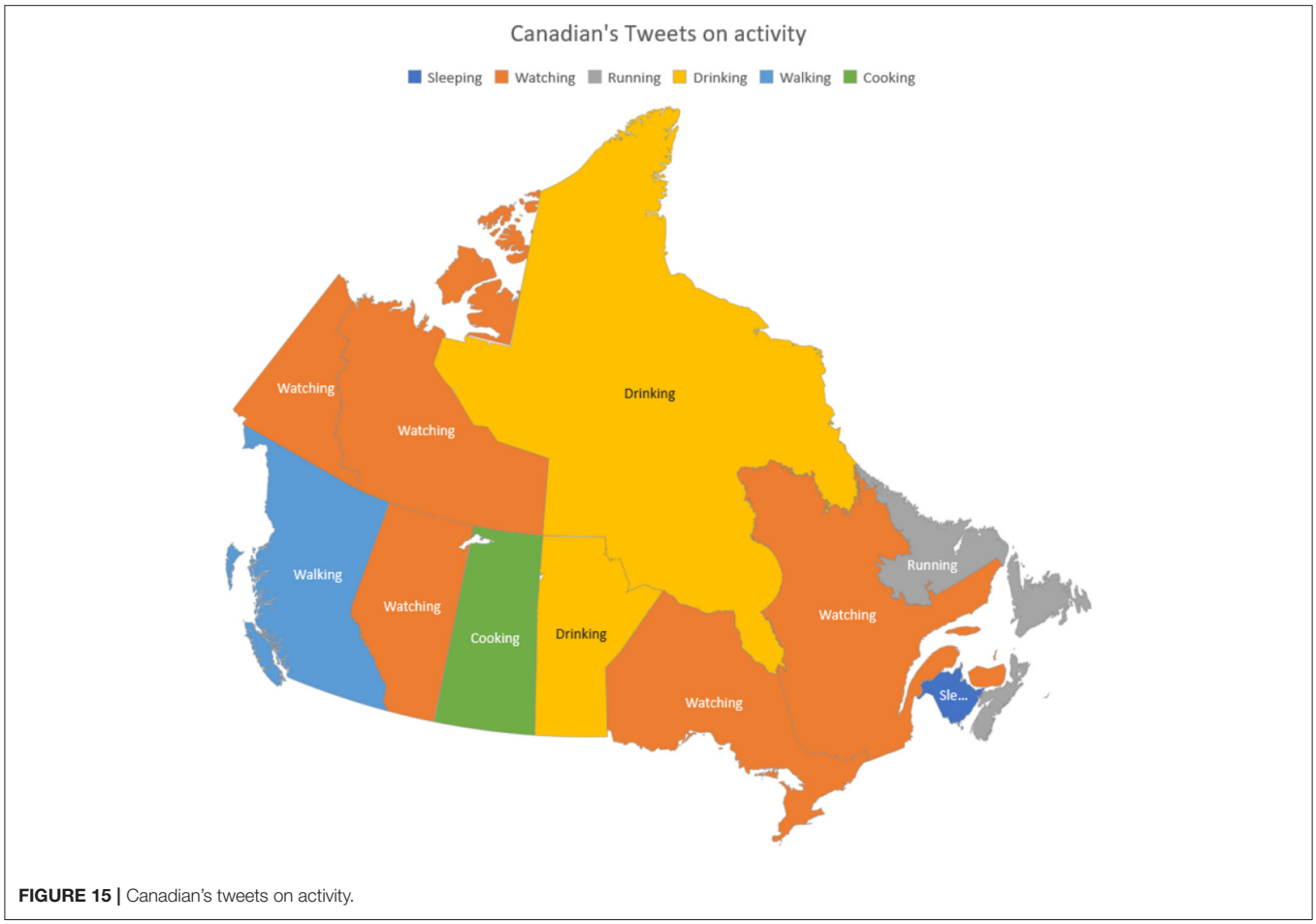

FIGURE 15 | Canadian's tweets on activity.

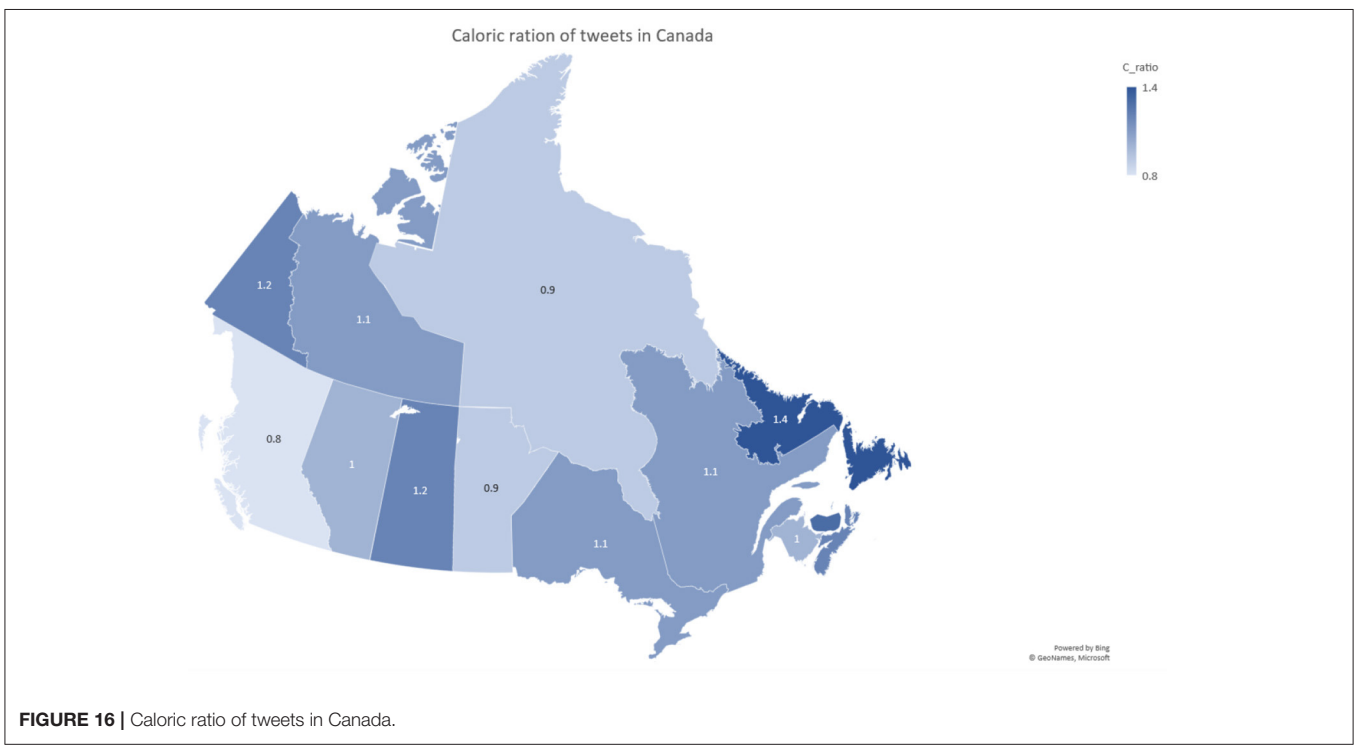

FIGURE 16 | Caloric ratio of tweets in Canada.

representation of real time. Here, many promising future works are possible such as a more dynamic way to calculate calories based on age, gender and work profile. Another limitation is that when our model looks at the tweet it only recognizes the food but it leaves out the quantity of said food. In example, our model will not be able to differentiate 1 apple from 10 apples.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. Authors interested in accessing the data can contact us at: https://www.datalab.science/.

AUTHOR CONTRIBUTIONS

NS developed the algorithm and wrote the first draft. GS and DS provided guidance to conduct the research. VM supervised the research and helped in all aspects of this research.

FUNDING

This research was partially funded by an NSERC Discovery Grant - RGPIN-2017-05377.

ACKNOWLEDGMENTS

The computing resources for the research are provided by the High Performance Computing (HPC) Lab and Department of Computer Science at Lakehead University, Canada. Authors are grateful to Mr. Chetan Mendhe for initially setting up the data collection stream, Dr. Salimur Choudhury for providing insight on the data analysis, and Mr. Malaravan Balachandran for reviewing and editing drafts.

REFERENCES

1. Social WA. *Global Digital Report 2018*. Erişim. Available online at: https://wearesocialcom/blog/2018/01/global-digital-report-2018 (2018).
2. Dodds PS, Harris KD, Kloumann IM, Bliss CA, Danforth CM. Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter. *PLoS ONE*. (2011) 6:e26752. doi: 10.1371/journal.pone.0026752
3. Disease Control Prevention. *Health-Related Quality of Life: Wellbeing Concepts* (2011).
4. Karami A, Dahl AA, Turner-McGrievy G, Kharrazi H, Shaw G Jr. Characterizing diabetes, diet, exercise, and obesity comments on Twitter. *Int J Inform Manage*. (2018) 38:1–6. doi: 10.1016/j.ijinfomgt.2017.08.002
5. Grover P, Kar AK, Davies G. Technology enabled Health—Insights from twitter analytics with a socio-technical perspective. *Int J Inform Manage*. (2018) 43:85–97. doi: 10.1016/j.ijinfomgt.2018.07.003
6. Lazer D, Kennedy R, King G, Vespignani A. The parable of Google Flu: traps in big data analysis. *Science*. (2014) 343:1203–5. doi: 10.1126/science.1248506
7. Paul MJ, Dredze M. You are what you Tweet: analyzing Twitter for public health. *Icwsm*. (2011) 20:265–72.
8. Hill S, Merchant R, Ungar L. Lessons learned about public health from online crowd surveillance. *Big Data*. (2013) 1:160–7. doi: 10.1089/big.2013.0020
9. Eichstaedt JC, Schwartz HA, Kern ML, Park G, Labarthe DR, Merchant RM, et al. Psychological language on Twitter predicts county-level heart disease mortality. *Psychol Sci*. (2015) 26:159–69. doi: 10.1177/0956797614557867
10. Culotta A. Estimating county health statistics with twitter. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Toronto, ON: ACM (2014). p. 1335–44.
11. Abbar S, Mejova Y, Weber I. You tweet what you eat: studying food consumption through twitter. In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. Seoul: ACM (2015). p. 3197–206.
12. Alajajian SE, Williams JR, Reagan AJ, Alajajian SC, Frank MR, Mitchell L, et al. The lexicocalorimeter: gauging public health through caloric input and output on social media. *PLoS ONE*. (2017) 12:e0168893. doi: 10.1371/journal.pone.0168893
13. Medvedyuk S, Ali A, Raphael D. Ideology, obesity and the social determinants of health: a critical analysis of the obesity and health relationship. *Crit Public Health*. (2018) 28:573–85. doi: 10.1080/09581596.2017.1356910
14. Diener E. Subjective well-being. *Psychol Bull*. (1984) 95:542.
15. Alajajian SE, Williams JR, Reagan JA, Alajajian SC, Frank RM, Mitchell L, et al. Available online at: https://figshare.com/articles/S1-dataset-tweetIDS_txt_zip/4530965/1 (2017).
16. McIntyre L, Jessiman-Perreault G, Mah CL, Godley J. A social network analysis of Canadian food insecurity policy actors. *Can J Diet Pract Res*. (2018) 79:60–6. doi: 10.3148/cjdpr-2017-034
17. Shah N, Willick D, Mago V. A framework for social media data analytics using Elasticsearch and Kibana. *Wireless Netw*. (2018) 1–9. doi: 10.1007/s11276-018-01896-2
18. Shah N. *Open Source Emoticons and Emoji Detection Library: Emot(stable v2.2)* (2018). Available online at: https://github.com/NeelShah18/emot (accessed February 28, 2019).
19. Batista R, Pottie K, Bouchard L, Ng E, Tanuseputro P, Tugwell P. Primary health care models addressing health equity for immigrants: a systematic scoping review. *J Immigr Minor Health*. (2018) 20:214–30. doi: 10.1007/s10903-016-0531-y
20. Williams JR. Boundary-based MWE segmentation with text partitioning. *arXiv [preprint] arXiv:160802025*. (2016). doi: 10.18653/v1/W17-4401
21. Mohan S, Thirumalai C, Srivastava G. Effective heart disease Prediction using hybrid machine learning techniques. *IEEE Access*. (2019) 7:81542–54. doi: 10.1109/ACCESS.2019.2923707
22. Mick P, Parfyonov M, Wittich W, Phillips N, Pichora-Fuller MK. Associations between sensory loss and social networks, participation, support, and loneliness: analysis of the Canadian longitudinal study on aging. *Can Fam Phys*. (2018) 64:e33–41.
23. Clarkson K, Srivastava G, Meawad F, Dwivedi AD. Where's @Waldo?: finding users on Twitter. In: *Artificial Intelligence and Soft Computing - 18th International Conference, ICAISC 2019*. Zakopane (2019). p. 338–49. doi: 10.1007/978-3-030-20915-5_31
24. Brook JR, Setton EM, Seed E, Shooshtari M, Doiron D. The Canadian Urban Environmental Health Research Consortium–a protocol for building a national environmental exposure data platform for integrated analyses of urban form and health. *BMC Public Health*. (2018) 18:114. doi: 10.1186/s12889-017-5001-5
25. Government of Canada Hig. *Obesity in Canadian Adults: It's About More Than Just Weight* (2017).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Artificial Intelligence

ORIGINAL RESEARCH
published: 31 March 2020
doi: 10.3389/frai.2020.00012

A Frequency-Domain Machine Learning Method for Dual-Calibrated fMRI Mapping of Oxygen Extraction Fraction (OEF) and Cerebral Metabolic Rate of Oxygen Consumption ($CMRO_2$)

Michael Germuska[1]*, *Hannah Louise Chandler*[1], *Thomas Okell*[2], *Fabrizio Fasano*[3], *Valentina Tomassini*[1,4,5], *Kevin Murphy*[1] and *Richard G. Wise*[1,5,6]

[1] Cardiff University Brain Research Imaging Centre (CUBRIC), Department of Psychology, Cardiff University, Cardiff, United Kingdom, [2] FMRIB, Nuffield Department of Clinical Neurosciences, Wellcome Centre for Integrative Neuroimaging, University of Oxford, Oxford, United Kingdom, [3] Siemens Healthcare Ltd., Camberley, United Kingdom, [4] Division of Psychological Medicine and Clinical Neurosciences, Cardiff University School of Medicine, Cardiff, United Kingdom, [5] Department of Neuroscience, Imaging and Clinical Sciences, "G. D'Annunzio University" of Chieti-Pescara, Chieti, Italy, [6] Institute for Advanced Biomedical Technologies, "G. D'Annunzio University" of Chieti-Pescara, Chieti, Italy

OPEN ACCESS

Edited by:
Shuihua Wang,
University of Leicester,
United Kingdom

Reviewed by:
Yi Chen,
Nanjing Normal University, China
Blaise Frederick,
Harvard Medical School,
United States

***Correspondence:**
Michael Germuska
germuskam@cardiff.ac.uk

Specialty section:
This article was submitted to Medicine and Public Health, a section of the journal Frontiers in Artificial Intelligence

Received: *31 May 2019*
Accepted: *09 March 2020*
Published: *31 March 2020*

Citation:
Germuska M, Chandler HL, Okell T, Fasano F, Tomassini V, Murphy K and Wise RG (2020) A Frequency-Domain Machine Learning Method for Dual-Calibrated fMRI Mapping of Oxygen Extraction Fraction (OEF) and Cerebral Metabolic Rate of Oxygen Consumption ($CMRO_2$). Front. Artif. Intell. 3:12. doi: 10.3389/frai.2020.00012

Magnetic resonance imaging (MRI) offers the possibility to non-invasively map the brain's metabolic oxygen consumption ($CMRO_2$), which is essential for understanding and monitoring neural function in both health and disease. However, in depth study of oxygen metabolism with MRI has so far been hindered by the lack of robust methods. One MRI method of mapping $CMRO_2$ is based on the simultaneous acquisition of cerebral blood flow (CBF) and blood oxygen level dependent (BOLD) weighted images during respiratory modulation of both oxygen and carbon dioxide. Although this dual-calibrated methodology has shown promise in the research setting, current analysis methods are unstable in the presence of noise and/or are computationally demanding. In this paper, we present a machine learning implementation for the multi-parametric assessment of dual-calibrated fMRI data. The proposed method aims to address the issues of stability, accuracy, and computational overhead, removing significant barriers to the investigation of oxygen metabolism with MRI. The method utilizes a time-frequency transformation of the acquired perfusion and BOLD-weighted data, from which appropriate feature vectors are selected for training of machine learning regressors. The implemented machine learning methods are chosen for their robustness to noise and their ability to map complex non-linear relationships (such as those that exist between BOLD signal weighting and blood oxygenation). An extremely randomized trees (ET) regressor is used to estimate resting blood flow and a multi-layer perceptron (MLP) is used to estimate $CMRO_2$ and the oxygen extraction fraction (OEF). Synthetic data with additive noise are used to train the regressors, with data simulated to cover a wide range of physiologically plausible parameters. The performance of the implemented analysis method is compared to published methods both in simulation and with *in-vivo* data ($n = 30$). The proposed

method is demonstrated to significantly reduce computation time, error, and proportional bias in both $CMRO_2$ and OEF estimates. The introduction of the proposed analysis pipeline has the potential to not only increase the detectability of metabolic difference between groups of subjects, but may also allow for single subject examinations within a clinical context.

Keywords: calibrated-fMRI, oxygen extraction fraction, $CMRO_2$, OEF, machine learning

INTRODUCTION

Under normal conditions the brain's energy needs are met via a continuous supply of oxygen and glucose for the local production of ATP via aerobic metabolism (Verweij et al., 2007). Any disruption of the supply of oxygen to the brain tissue can have significant consequences (Safar, 1988), and impaired cerebral oxygen metabolism is associated with a wide variety of neurological conditions (Frackowiak et al., 1988; Ishii et al., 1996; Miles and Williams, 2008). Therefore, monitoring and mapping the brain's consumption of oxygen is vital for understanding the diseases and mechanisms by which the metabolic consumption of oxygen may be affected. The cerebral metabolic rate of oxygen consumption ($CMRO_2$) has traditionally been measured with positron emission tomography (Frackowiak et al., 1980). However, this method has some substantial limitations including the use of ionizing radiation and the need for local production of 15-oxygen labeled tracers. Due to these limitations there is great interest in developing alternative, non-invasive, methods of mapping $CMRO_2$. One promising technique of non-invasively mapping $CMRO_2$ is the so-called dual-calibrated fMRI (dc-fMRI) method (Bulte et al., 2012; Gauthier et al., 2012). This method is finding growing adoption in the research setting, and has already been applied in Alzheimer's disease (Lajoie et al., 2017), carotid artery occlusion (De Vis et al., 2015), and studies of pharmacological modulation (Merola et al., 2017). For a review of the method and details on the its practical application please see Germuska and Wise (2019). Despite the promise shown by this technique, the reported between-session repeatability is relatively low (Merola et al., 2018) and improvements in the data acquisition and/or analysis are required if individualized assessment is to be made possible.

One of the key difficulties in analyzing dual-calibrated fMRI data is noise propagation through the analysis pipeline, which leads to unstable parameter estimates. We have previously presented regularized non-linear least squares fitting approaches that utilize prior physiological knowledge to produce more robust parameter estimates (Germuska et al., 2016, 2019). Even though such regularization reduces the mean square error it does so by trading off a reduction in variance with an increase in bias. An alternative approach to reduce the prediction error is the use of noise insensitive machine learning regression methods. Decision tree based regression methods, for example random forest (Breiman, 2001) and extremely randomized trees (Geurts et al., 2006), are robust to both output (Breiman, 2001; Geurts et al., 2006) and input noise (Yue et al., 2018) and are able to capture non-linear relationships between input features and target parameters. This noise immunity is likely due to the randomization included in the choices of features at splitting nodes (random forest) and cut-points (extremely randomized trees), which improve the generalizability of the regressors. For non-linear mappings with a high degree of complexity artificial neural networks, such as the multi-layer perceptron (MLP), a feedforward network with multiple hidden layers, offer a machine learning method that is inherently robust to noise (Bernier et al., 1999). In this paper we present an analysis pipeline comprised of an extremely randomized trees regressor and a MLP, cascaded to infer resting CBF and $CMRO_2$ from dual-calibrated fMRI data. A frequency-domain representation of simulated MRI data with the additive noise is used to train each of the regressors. Simulated data has the advantage over *in-vivo* data in this application as it allows a balanced dataset to be generated that covers a broad range of physiological variation. Such a dataset is essential to avoid bias in parameter estimation and to provide generalizability across groups and diseases. A frequency-domain representation is chosen as it allows for convenient dimensionality reduction, with most of the information of interest encoded at low temporal frequencies, and takes advantage of the superior ability of artificial neural networks to learn discriminative features from frequency-domain representation of a signal compared to a time-domain representation (Hertel et al., 2016). The performance of the proposed machine learning (ML) implementation is compared to an existing regularized non-linear least squares (rNLS) method (Germuska et al., 2019) both in simulation and in data acquired from a cohort of 30 healthy volunteers. We hypothesized that the machine learning approach would be able to achieve comparable or reduced prediction error with significantly reduced bias and computational overhead.

MRI DATA ACQUISITION

Thirty healthy volunteers (16 males, mean age 32.53 $\pm$ 6.06 years) were recruited to the study, see **Supplementary Table 1** for demographic data. The local ethics committee approved the study and written informed consent was obtained from each participant. Blood samples were drawn via a finger prick prior to scanning and were analyzed with the HemoCue Hb 301 System (HemoCue, Ängelholm, Sweden) to calculate the systemic [Hb] value for each participant. All data was acquired using a Siemens MAGNETOM Prisma (Siemens Healthcare GmbH, Erlangen) 3T clinical scanner with a 32-channel receiver head coil (Siemens Healthcare GmbH, Erlangen). The acquisition protocol was as previously described (Germuska et al., 2019). Briefly, an 18-min dual-excitation pseudo-continuous arterial

TABLE 1 | Range of physiological parameters used in the dc-fMRI data simulations for training of the machine learning regressors.

OEF	CBF (ml/100g/min)	[Hb] (g/dL)	Mean capillary transit time (CBV_{cap}/CBF, s)	$PminO_2$ (mmHg)	Cerebral vascular reactivity (% CBF/mmHg CO_2)	K
0.05–0.75	1–250	10–18	0.25–4.0	0–30	1–7	0.01–0.25

spin labeling (pCASL) and BOLD-weighted acquisition was acquired during modulation of inspired oxygen and carbon dioxide. Gas modulation was performed according to a protocol previously proposed by our lab (Germuska et al., 2016), and end-tidal monitoring was performed throughout the acquisition from the volunteer's facemask using a rapidly responding gas analyzer (PowerLab®, ADInstruments, Sydney, Australia). The prototype pCASL sequence (Germuska et al., 2019) parameters were as follows: post-labeling delay and label duration 1.5 s, EPI readout with GRAPPA acceleration (factor = 3), $TE_1 = 10$ ms, $TE_2 = 30$ ms, TR = 4.4 s, 3.4 × 3.4 mm in-plane resolution, and 15 (7 mm) slices with 20% slice gap.

SYNTHETIC MRI DATA GENERATION

Synthetic data was simulated to match the 18-min *in-vivo* acquisition protocol using standard physiological models for the change in BOLD signal (Bulte et al., 2012; Gauthier and Hoge, 2013; Wise et al., 2013), as summarized by Equation (1).

$$\frac{\Delta BOLD}{BOLD_0} = \mathrm{M}\left\{1 - \left(\frac{CBF}{CBF_0}\right)^{\alpha}\left(\frac{1 - \frac{CaO_2 - CMRO_2/CBF}{\varphi[Hb]}}{1 - \frac{CaO_{2,0} - CMRO_{2,0}/CBF_0}{\varphi[Hb]}}\right)^{\beta}\right\} \quad (1)$$

Where, $BOLD/BOLD_0$ is the fractional change in BOLD signal due to a change in arterial oxygen content (CaO_2) or CBF due to either a hyperoxic or hypercapnic respiratory stimulus. M is a lumped parameter that is equal to $K \cdot TE \cdot \left((1 - SvO_2) \cdot [Hb]\right)^{\beta}$. Where K is a scaling factor dependent on the field strength, resting venous blood volume, tissue structure, and water diffusion effects in the extravascular space. [Hb] is the blood hemoglobin concentration and SvO_2 is the venous oxygen saturation. ϕ is the oxygen binding capacity for Hb (1.34 ml/g), α is the Grubb exponent that couples blood volume and blood flow changes, and β is a field strength dependent constant that summarizes the non-linear effects associated with the tissue structure and water diffusion effects. The values of α and β were fixed to the optimized values (0.06 and 1) found by Merola et al. (2016), which minimize the error in OEF estimates over a range of vascular physiology. The subscript 0 represents the baseline or resting state. The hyperoxic and hypercapnic stimuli are assumed to be iso-metabolic, so $CMRO_2 = CMRO_{2,0}$.

The arterial spin labeling signal was modeled according to the simplified pCASL kinetic model (Alsop et al., 2015), and physiological constraints on baseline parameters were applied according to a simple model of oxygen exchange (Gjedde, 2002; Hayashi et al., 2003) (Equation 2).

$$CMRO_{2,0} = \mathrm{D}\left[P_{50}\sqrt[h]{\frac{2}{OEF_0} - 1} - P_{\min O_2}\right] \quad (2)$$

Where D is the effective oxygen diffusivity of the capillary network and can be expressed as a product of the effective oxygen permeability and the capillary blood volume, $D = \kappa \cdot CBV_{cap}$. P_{50} is the blood oxygen tension at which hemoglobin is 50% saturated (26 mmHg), h is the Hill coefficient (2.8) and $P_{\min O_2}$ is the minimum oxygen tension at the mitochondria [which is thought to be negligible in healthy tissue (Gjedde, 2002)]. In the modeling we assume a fixed value for κ of 3 μmol/mmHg/ml/min, corresponding to a typical diffusivity of 3 (Mintun et al., 2001) to 4 μmol/100 g/mmHg/min (Vafaee and Gjedde, 2000) for CBV_{cap} = 1–1.33 ml/100 g. The physiological parameter space encompasses a wide range of plausible physiology including both healthy and dysfunctional brain tissue, and is summarized in **Table 1**. A summary of MRI abbreviations and all model parameters used in the simulations is given in **Table 2**.

The partial pressure of arterial oxygen (PaO_2) and change in carbon dioxide ($\Delta PaCO_2$) were modeled to match the range of end-tidal recordings acquired from healthy volunteers. The baseline PaO_2 had a range of 90–120 mmHg, ΔPaO_2 was 200–300 mmHg, and $\Delta PaCO_2$ was set to 8–12 mmHg. Rectangular stimulus blocks were convolved with a gamma density function with shape parameter 0.5–2.5 to account for the variation in biological rise and fall times of the hyperoxic and hypercapnic stimuli. Drift in $\Delta PaCO_2$, which was observed in some subjects, was included by adding a bandpass filtered noise signal (fourth order IIR filter, lowcut/highcut = 0.005/0.05 of the Nyquist frequency). Change in the arterial blood longitudinal relaxation rate due to dissolved oxygen was included in pCASL calculations as per (Germuska et al., 2019). Noise (BOLD tSNR = 90, pCASL tSNR = 3 for CBF = 60 ml/100 g/min) was added to simulated BOLD and pCASL time series. The pCASL noise was bandpass filtered (fourth order IIR filter, lowcut/highcut = 0.05/0.8 of the Nyquist frequency) and the BOLD noise was lowpass filtered (first order IIR filter, highcut = 0.5 of the Nyquist frequency) to match the noise characteristics of the *in-vivo* data. In addition, the BOLD timeseries data was highpass filtered with a 320 s cut-off using the filter implementation in FSL (Jenkinson et al., 2012), which is routinely used for de-trending fMRI data. **Figure 1** shows 50 randomly generated pCASL and BOLD timeseries overlaid with the temporal mean to demonstrate the typical output of the simulations. Please note that the pCASL timeseries are divided by the equilibrium magnetization of arterial blood ($M0_{blood}$), and the baseline signal has been set to zero for display purposes.

TABLE 2 | Summary of model parameters and abbreviation used in the dc-fMRI data simulations and their definitions.

Variable/ abbreviation	Expression (units)
OEF	Oxygen Extraction Fraction (dimensionless)
$CMRO_2$	Cerebral Metabolic Rate of Oxygen consumption (μmol/100 g/min)
CBF	Cerebral Blood Flow (ml/100 g/min)
ϕ	Oxygen binding capacity of hemoglobin (1.34 ml/g)
[Hb]	Hemoglobin concentration (g/dL)
CaO_2	Arterial oxygen content (ml/ml)
PaO_2	Arterial oxygen tension (mmHg)
SaO_2	Arterial oxygen saturation (dimensionless)
SvO_2	Venous oxygen saturation (dimensionless)
α	Grubb exponent
β	Venous morphology/deoxy-hemoglobin—BOLD exponent
BOLD	Blood Oxygenation Level Dependent signal
ASL	Arterial Spin Labeling
$M0_{blood}$	Arterial blood MRI signal equilibrium magnetization (dimensionless)
PLD	ASL post-label delay time (1.0–3.0 s)
M	Maximum possible BOLD signal (BOLD calibration parameter)
K	BOLD scaling factor = $M/([Hb] \times (1 - SvO_2))^{\beta}$
D	Effective oxygen diffusivity of the capillary network (μmol/100 g/mmHg/min)
CBV_{cap}	Capillary blood volume (ml/100 g)
$PminO_2$	Minimum oxygen partial pressure at the mitochondria (mmHg)
h	Hill coefficient (2.8)
κ	Effective permeability of capillary endothelium and brain tissue (μmol/mmHg/ml/min)

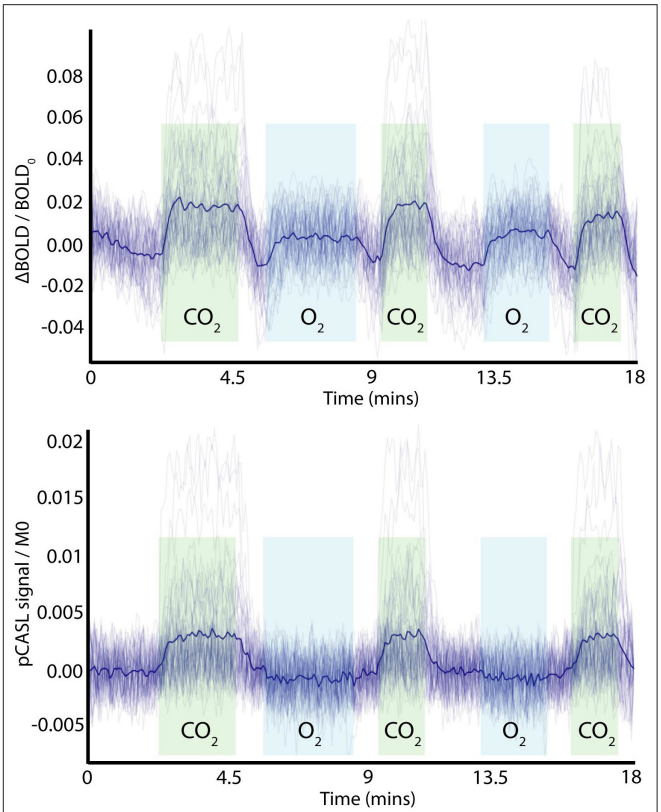

FIGURE 1 | Example of simulated time-domain data (BOLD and ASL) with added noise and variation in physiological parameters, showing periods of hypercapnic (green) and hyperoxic (light blue) stimuli. The dark blue line represents the mean time-course over the example time series. Note the pCASL signal is normalized by the equilibrium magnetization of arterial blood (MO) and has had the baseline signal subtracted for display purposes.

METHODS

A schematic diagram describing the analysis/training pipeline is shown in **Figure 2**. ASL and BOLD timeseries data, either simulated (as described in section Synthetic MRI Data Generation) or *in-vivo* data, are Fourier transformed into magnitude and phase data. This frequency domain data is then truncated after the first 15 data points (low pass filtered) and combined with physiological recordings and sequence parameters to create a feature vector for model training/prediction (if *in-vivo* data is being analyzed). Parameter estimation is carried out in a two-stage process; first the resting blood flow (CBF_0) is estimated, and then rate of oxygen consumption.

Truncation of the frequency domain data removes high-frequency content that is unrelated to either the hyperoxic or hypercapnic respiratory modulations and thus removes noise from the training data. The resting blood flow is estimated separately from the rate of oxygen consumption to reduce the complexity of the required mapping between the MRI data and the target parameters. Additionally, the use of extremely randomized trees (ET) regression rather than an artificial neural network at this stage in the pipeline takes full advantage of the noise immunity of decision tree based methods (Yue et al., 2018) and reduces the potential of overfitting. The inclusion of the post-label delay in the feature vector is necessary to incorporate an implicit slice timing correction for CBF_0 calculation, while the blood oxygenation parameters ([Hb], ΔPaO_2, $SaO_{2,0}$, $CaO_{2,0}$) are included here due to the influence of dissolved oxygen on the longitudinal relaxation rate of arterial blood. In total each feature vector that is input into the ET regressor consists of 65 entries.

The result of the ET regression is then incorporated into the feature vector (now 66 entries) and input into an ensemble of MLPs to predict $CMRO_{2,0}/CaO_{2,0}$, from which $CMRO_{2,0}$ and OEF_0 can be calculated ($CMRO_2/CaO_2 = OEF \times CBF$ via the Fick principle). The blood oxygenation parameters in this case not only inform on the relaxation rate of arterial blood, but also link the CBF and BOLD signal changes to the underlying metabolic parameters as described by Equation (1). In practice each MLP in the ensemble is trained individually, with the average of their predictions being used for inference when deployed for the analysis of *in-vivo* data.

The ET regressor and MLP were implemented in Scikit learn (Pedregosa et al., 2011). The extremely randomized trees regressor was trained with the following options, number of estimators = 50, bootstrap = True, and out-of-bag samples were used to estimate the R^2 on unseen data. A total of 50,000

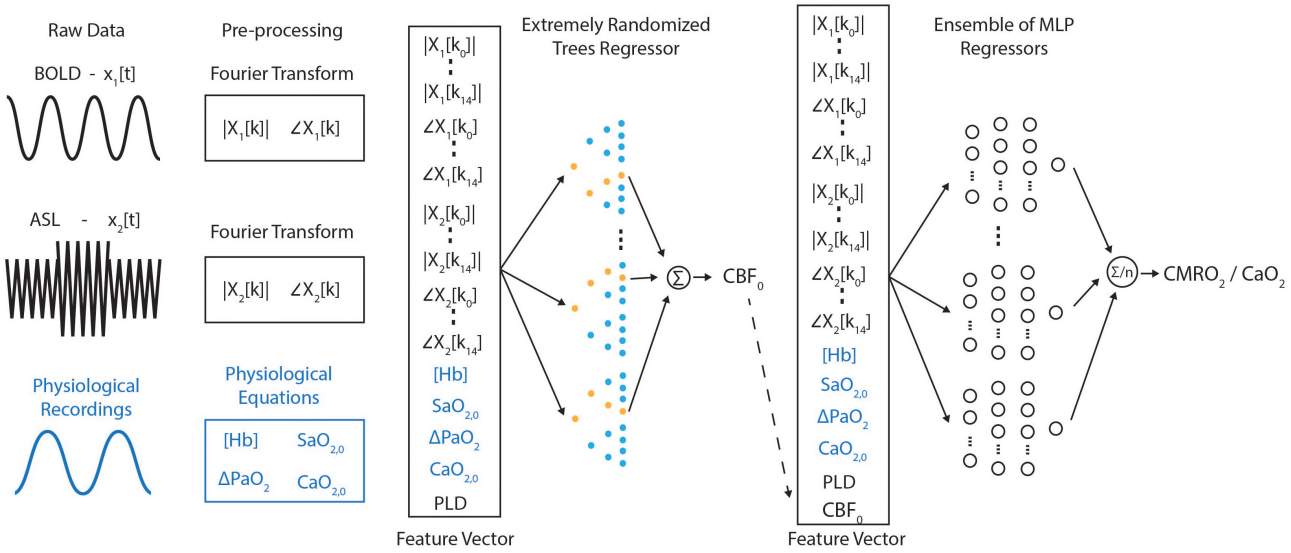

FIGURE 2 | Schematic diagram of the frequency-domain machine learning pipeline. Raw data is pre-processed prior to the construction of a feature vector. This initial feature vector is used to estimate baseline perfusion. The perfusion estimate is then included in the feature vector fed into an ensemble of multilayer perceptron networks used to estimate the resting rate of oxygen metabolism.

simulations were used for training. The MLP network has two-hidden layers and 50 nodes in each layer. The activation function for each node was chosen to be a rectified linear unit (ReLU). The ADAM solver was used for training with 1×10^6 simulated feature vectors and 10% of the data were used for early stopping. Data simulation and training was repeated 40 times to create an ensemble of MLP networks to further reduce the uncertainty in parameter estimates (Sollich and Krogh, 1996).

The validation score for the extremely randomized trees regressor for predicting resting cerebral blood flow was 0.997, slightly greater than the results obtained for a random forest implementation (0.961). The validation score for the MLP estimation of $CMRO_{2,0}/CaO_{2,0}$ were 0.923 ± 0.002. Training of the MLP network was also undertaken while eliminating key elements of the simulation or feature vectors to see how this affected the performance of the MLP. When BOLD data was excluded from the feature vector the validation score dropped to 0.577. Excluding the CO_2 and O_2 stimuli (but including the BOLD data) reduced the validation scores to 0.63 and 0.71, respectively.

A further 5,000 simulated datasets (with OEF restricted to 0.15–0.65, all other parameters as in **Table 1**) were constructed to compare the performance of the proposed machine learning implementation with a previously implemented regularized non-linear least squares fitting method (Germuska et al., 2019). Each method was compared to the simulated data using a robust regression method (bisquare) in terms of the RMS error and proportional bias. A bisquare cost function was used for the regression to reduce the influence of outliers and allow a robust estimate of the proportional bias. The rNLS fitting was implemented with regularization applied to the resting OEF and the effective oxygen diffusivity (D), as previously described. The relative weighting between OEF and diffusivity regularization was maintained constant, as per the optimization in Germuska et al. (2019). However, the total weighting was varied to assess the impact on OEF and $CMRO_2$ error and proportional bias (slope of the simulated parameter values plotted against the parameter estimates).

RESULTS

Simulations

Analysis of the simulated data demonstrated a substantial reduction in the RMS error of machine learning OEF estimates compared to rNLS estimates. The bisquare RMS error was 0.047 when using the mean prediction from the 40 MLP networks, and 0.055 for a randomly chosen MLP network. The rNLS approach produced a minimum bisquare RMS error of 0.094. The ML approach displayed negligible proportional bias in OEF estimates (slope of true vs. estimated values = 0.982), whereas rNLS estimates had variable levels of bias depending on the level of regularization, see **Figure 3A** for a summary of the results. As expected from the OEF results, ML estimates of $CMRO_2$ also had significantly reduced error and bias compared to the rNLS implementation. The proportional bias for the ML implementation was 0.977 compared to a minimum bias of 0.913 for the rNLS method. The bisquare RMS error in $CMRO_2$ estimates for the ML implementation was 20.3 μmol/100 g/min (22.6 for an individual MLP network) whereas the error for rNLS estimates ranged from 29.6 to 52.4 μmol/100 g/min depending on the level of bias (with greater bias coinciding with reduced error) (see **Figure 3B**).

Training of the MLP with reduced feature vectors (excluding the BOLD data) or limited respiratory stimuli (excluding either

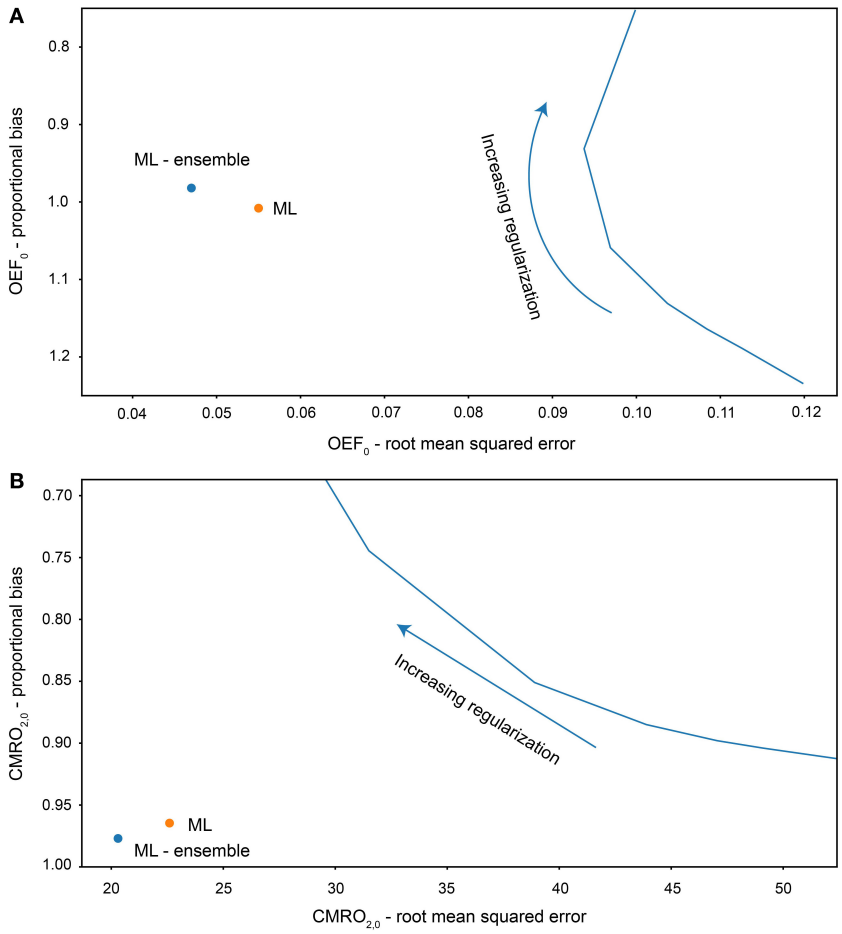

FIGURE 3 | Root mean squared error and proportional bias in OEF_0 **(A)** and $CMRO_{2,0}$ **(B)** estimates for each analysis method fitting to simulated data (5,000 simulations). Solid blue line plots the error and bias for increasing regularization weighting for the regularized non-linear least squares analysis.

CO_2 or O_2 modulation) highlights the importance of each signal and stimulus in estimate the rate of oxygen consumption. As expected, removing the BOLD signal resulted in a significant reduction in the network's ability to estimate $CMRO_2$ (validation R^2 reduced from 0.923 for the full model to 0.58). In this instance there should be no information relating to OEF in the feature vector and so the inference is based solely on the correlation between baseline flow and $CMRO_2$ in the simulated data. Adding the BOLD data back in but with only an O_2 stimulus does little to improve the performance of the network ($R^2 = 0.63$). This is not unexpected as the hyperoxic BOLD signal is largely related to venous blood volume (Blockley et al., 2013) with little influence from OEF. Perhaps unexpectedly, including the CO_2 stimulus but not the O_2 stimulus significantly improves the ability of the network to infer resting $CMRO_2$ ($R^2 = 0.71$). While this is still significantly worse than the full model, it suggests that some quantitative metabolic information may be extracted from hypercapnic calibration studies that are normally employed to estimate relative changes in $CMRO_2$ (Hoge, 2012). Additionally, such results suggest that the simulation framework could be utilized to optimize data acquisition by designing respiratory stimuli that maximize the performance of the ML implementation, and that such respiratory paradigms may be different compared to those for standard analysis methods (which are unable to infer resting $CMRO_2$ information from a hypercapnic calibration experiment).

In-vivo

Due to the limited availability and technical challenges associated with acquiring 15-oxygen PET data for $CMRO_2$ mapping (the gold standard approach) it is difficult to directly validate the *in-vivo* results obtained in this study. However, a number of fundamental relationships between resting physiological parameters have consistently been observed across groups of healthy individuals. Here we compare these observed relationships against the acquired data to infer the relative error and bias for each analysis method. One of the most frequently reported relationships in the healthy human brain is that resting blood flow is linearly correlated with resting oxygen metabolism (Scheinberg and Stead, 1949; Lebrun-Grandie et al., 1983; Leenders et al., 1990; Coles et al., 2006; Powers et al., 2011). Additionally, PET data suggests that the OEF should

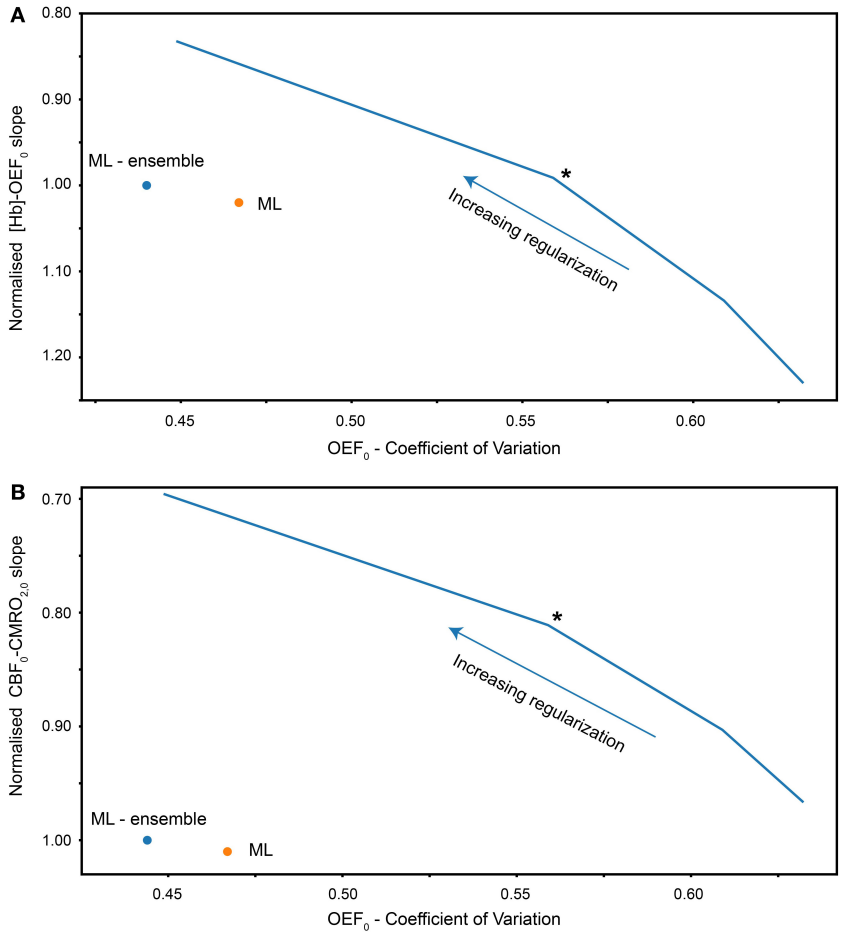

FIGURE 4 | (A) Coefficient of variation of gray matter OEF_0 estimates vs. slope of [Hb]-OEF_0 relationship for each analysis method (rNLS fitting evaluated with increasing levies of regularization). The [Hb]-OEF_0 slope has been normalized by the ML ensemble estimate of the [Hb]-OEF_0 slope. **(B)** Coefficient of variation of gray matter OEF_0 estimates vs. the slope of the CBF-$CMRO_2$ relationship, normalized by the ML (ensemble) estimate of the CBF-$CMRO_2$ slope. Solid blue line plots the coefficient of variation against the slope for increasing regularization weighting for regularized non-linear least squares analysis. The asterisk indicates the chosen level of regularization for subsequent analysis/comparisons.

be approximately uniform across the cerebral gray matter (e.g., Hyder et al., 2016). Thus, we can use the coefficient of variation (COV) of gray matter OEF estimates as an indicator of parameter error, and examine the variation in the slope of the CBF-$CMRO_2$ relationship to infer the proportional bias or sensitivity to physiological variation of $CMRO_2$ estimates.

As in the simulation experiments we investigated the *in-vivo* analysis for varying levels of regularization in the rNLS analysis and compare this to the ML results. **Figure 4B** plots the COV in OEF estimates for increasing levels of regularization against the slope of the CBF-$CMRO_2$ regression (normalized by the slope of the ML estimate). As predicted by the simulations, the slopes of the ML estimates and the rNLS estimates are similar when little regularization is applied, with the slope of the rNLS estimates slightly reduced compared to the ML approach. As more regularization is applied the COV of OEF estimates is reduced and the slope between CBF and $CMRO_2$ decreases, clearly demonstrating the trade-off between variance and bias. Again, as predicted by the simulations, the COV in ML estimates is significantly less than COV in rNLS estimates for a similar CBF-$CMRO_2$ slope.

To investigate the bias in OEF estimates we take advantage of another physiological relationship reported in the literature; cerebral oxygen extraction is inversely related to [Hb] (Ibaraki et al., 2010) and the closely related parameter Hct (Morris et al., 2018). Taking the same approach as before we observe *in-vivo* results that closely match predictions from the simulation (see **Figure 4A**). As in the simulations, the slope in the [Hb]-OEF relationship is similar between the ML method and rNLS approach for a moderate amount of regularization. However, the slope is substantially increased when using minimal regularization, and reduced when applying strong regularization.

Figure 5 shows scatter plots of the gray matter CBF-$CMRO_2$ and [Hb]-OEF relationships observed with the ML and rNLS methods across the 30 healthy volunteers studied. The rNLS results are shown for a single level of regularization, where the

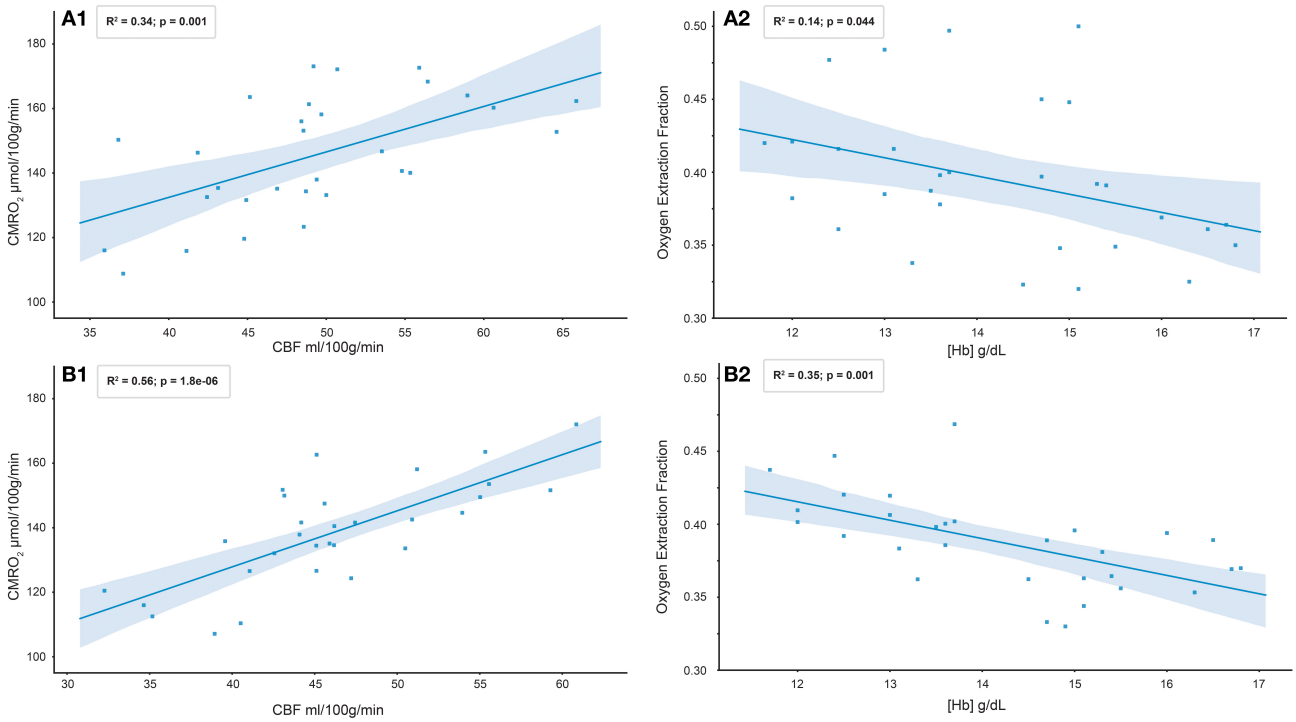

FIGURE 5 | Scatter plots of gray matter C8F-CMRO$_2$ and [Hb]-OEF relationships observed with rNLS (A1 and A2) and ML ensemble (B1 and B2) methods across 30 healthy volunteers.

slope of the [Hb]-OEF relationship most closely matches that of the ML analysis (see **Figure 4**). The coefficient of determination is greater for the ML approach for each relationship, with R^2 values of 0.56 and 0.35 for the CBF-CMRO$_2$ and [Hb]-OEF relationships, compared to 0.34 and 0.14 for the rNLS approach ($p < 0.05$ for all correlations).

Table 3 reports the results of a bivariate regression of OEF against [Hb] and CBF for both analysis methods. The slopes of the relationship between OEF and [Hb] are similar to that reported in healthy subjects by (Ibaraki et al., 2010), −1.75 Hb (g/dL). As per Ibaraki et al. the relationship between CBF and OEF did not reach significance ($p = 0.44$) for the ML approach, however a significant negative correlation was observed in the rNLS analysis ($p = 0.005$). A univariate analysis of CMRO$_{2,0}$ against CBF$_0$ is consistent with that observed in healthy controls by (Powers et al., 2011) ($\beta1 = 0.2$) for both analysis methods, $\beta1 = 0.32$ ($p < 0.001$) and $\beta1 = 0.24$ ($p < 0.001$) for the ML and rNLS approaches respectively.

Figure 6 shows a comparison between CBF$_0$, OEF$_0$, and CMRO$_{2,0}$ parameter maps calculated with the ML method (single MLP network and ensemble of 40 networks) and the rNLS method. The image shows seven slices from a single subject, which have been interpolated for display using cubic b-spline interpolation (Ruijters and Thevenaz, 2012) using FSLeyes (10.5281/zenodo.1470761). As expected OEF$_0$ is not well-estimated in the white matter, due to the T$_1$ decay of the arterial spin labeling signal and the longer arrival time of white matter blood. Across gray matter containing voxels maps of OEF$_0$ calculated with the ML methods are more uniform than those calculated with the rNLS approach, with the ensemble approach visibly outperforming the singe network MLP estimates. These observations are consistent with the results of the simulations and the gray matter COV observed for *in-vivo* OEF$_0$ estimates. However, it is also apparent from the images that each method demonstrates sensitivity to regional susceptibility effects. For example, in the pre-frontal cortex and inferior temporal lobes the images show greater variability in OEF$_0$ estimates, with regions of both over and under-estimation apparent. This instability is likely due to reduced BOLD SNR in these locations and alteration of the susceptibility of air in and around the nasal cavity and paranasal sinuses due to modulation of the inspired oxygen content during data acquisition. It is clear that the ML estimates, in particular those made from the ensemble of MLPs, are more robust to such regional susceptibility effects.

TABLE 3 | Results of a bivariate regression of OEF$_0$ against CBF$_0$ and [Hb] for 30 healthy volunteers analyzed with the ML (ensemble of MLPs) and rNLS fitting methods.

Predictor	ML β1 (*p* value)	rNLS β1 (*p* value)
[Hb]	−1.42 (0.001)	−2.23 (0.001)
CBF	−0.07 (0.44)	−0.37 (0.005)
Intercept	61.95 (<0.001)	89.48 (<0.001)

The *in-vivo* analysis also highlights the improvement in computational efficiency of the proposed method. The rNLS

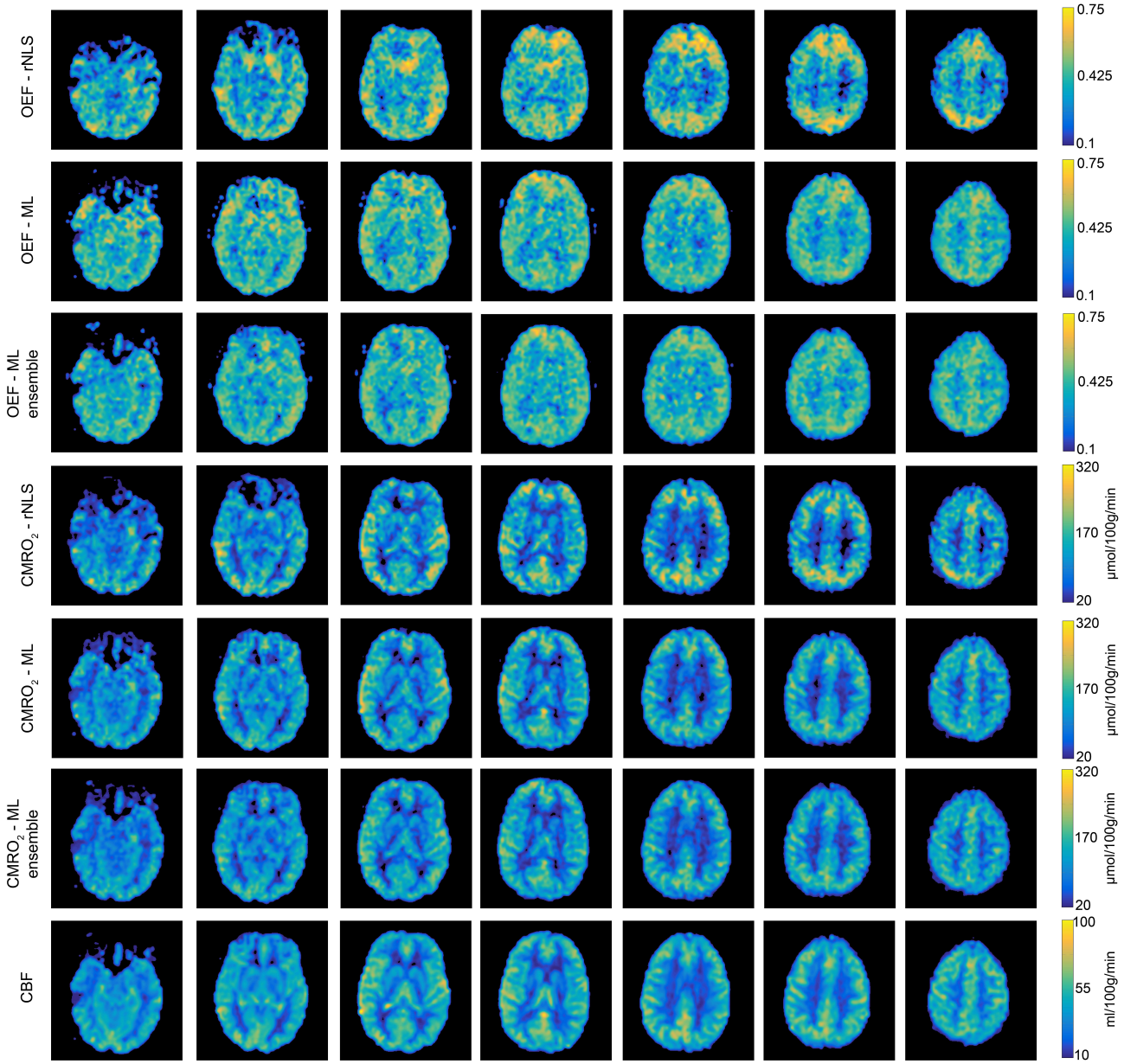

FIGURE 6 | Example parameter maps (CBF_0, OEF_0, and $CMRO_{2,0}$) from a single subject for each analysis method. Machine learning estimates of OEF_0 are more uniform than regularized non-linear least squares estimates. Using an ensemble of MLP networks further reduces the spatial variation in OEF_0 estimates.

approach took ∼20 min to analyze a complete dataset on a standard laptop (2.8 Ghz Intel Core i7, 16GB memory), while the ML approach was able to complete the same analysis in ∼10–20 s (depending on the number of networks in the ensemble of MLP regressors).

DISCUSSION AND CONCLUSIONS

Instability in parameter estimates made using noisy *in-vivo* data may be reduced by incorporating prior knowledge of physiological parameters (e.g., Chappell et al., 2010; Frau-Pascual et al., 2014; Mesejo et al., 2015; Germuska et al., 2016). Previous investigation of such methods (Germuska et al., 2016) suggests that they are an effective means to increase the robustness of $CMRO_2$ estimates made with dc-fMRI. However, these methods are computationally expensive and must necessarily make a trade off between parameter uncertainty and parameter sensitivity. Thus, they are not well-suited to high throughput or rapid data analysis and care must be taken when using such methods not to unduly bias parameter estimates toward the priors. In the work presented here we take a different approach by training a machine learning implementation that is robust to input noise.

Given an appropriately selected (or generated) training dataset, a well-implemented solution will be unbiased, robust, and have a low computational overhead.

Computer modeling suggests that the proposed method outperforms previous analysis methods both in terms of uncertainty and bias. *In-vivo* data supports the predicted improvement in uncertainty with a significant reduction in the COV of gray matter OEF_0 estimates when compared to a regularized non-linear least squares fitting of the data. Additionally, agreement was found between the predicted behaviors of each method and their associated biases when compared to reported physiological relationships. Qualitatively, the *in-vivo* parameter maps suggest that the ML approach, especially when paired with an ensemble implementation, is more robust to physiological noise; producing physiologically plausible parameter estimates in challenging brain regions, e.g., near the frontal sinuses. Such physiological noise was not modeled in the training data so it is perhaps unexpected that the ML method is robust to these noise sources. However, it is plausible that the discriminative features identified from the frequency-domain representation of the data during training are less sensitive to these regional susceptibility changes than a traditional time-domain fit of the data. It is possible that this aspect of the ML approach could be enhanced by extending the training data to include such regional susceptibility changes, either on their own or in combination with a spatially informed approach to data fitting.

The use of an ensemble of MLP networks reduced parameter uncertainty in simulation and reduced the coefficient of variation in gray matter OEF_0 estimates *in-vivo*, demonstrating its utility in this application. However, it is anticipated that enforcing network diversity during training could make further improvements in performance. As it is has previously been demonstrated that, in the presence of noise, the performance of an ensemble of networks can always be improved by explicitly encouraging diversity during training (Reeve and Brown, 2018).

The machine learning implementation presented here employs a combination of proven signal processing (time-frequency transformation) and machine learning methods (decision trees and fully connected artificial neural networks) that have been shown to select appropriate features for learning and are robust to input noise. The proposed analysis pipeline demonstrates an improvement in both the accuracy and precision in parameter estimates compared to published methods, and is appropriate for the study of both healthy volunteers and in clinical investigations. However, there are still many avenues that could be explored both in terms of signal processing and machine-learning. For example time domain data could be converted to 2D time-frequency representations, such as a spectrogram, or into spectrogram-like representations using wavelet transforms (for increased time resolution). This type of pre-processing would open the door to the application of 2D convolution neural networks (CNN) that have been so successfully applied in the domain of image processing. It is possible that the application of such approaches could further improve the performance of machine learning when analyzing dc-fMRI data. However, a thorough investigation of all available machine learning methods and associated pre-conditioning of the data is beyond the scope of the current study, which focuses instead on the realization of a practical solution by combining well-proven techniques for the analysis of signal data.

All *in-vivo* analysis in this manuscript is performed in the absence of spatial smoothing, which is often employed to improve statistical estimates made from fMRI data (Friston et al., 1995). We chose not to employ spatial smoothing in this analysis for two principle reasons: first any such spatial filtering implies a prior assumption regarding the spatial extent of any variation (Rosenfeld and Kak, 1982), and can thus lead to unwanted loss of sensitivity to physiological variation; second we did not want to increase the potential contamination of gray matter voxels with non-tissue signals, such as CSF or macrovessels (both of which are not included in the underlying signal model). The current study does not make any direct comparison between smoothed and unsmoothed analysis pipelines, however the presented method clearly avoids any possible smoothing artifacts that might otherwise bias the analysis.

A limitation of the proposed method is the need to train new regressors for a given gas paradigm and set of acquisition parameters, e.g., arterial spin labeling tagging duration, repetition time and duration of the acquisition. In addition, there is a requirement that the *in-vivo* gas manipulation does not deviate significantly from the range of simulated designs. While it is a relatively straightforward process to retrain the regressors with a new set of parameters, to match the local acquisition protocol, the scope of the method could be increased if individualized gas traces could be incorporated into the training data; allowing a single pre-trained implementation to be applied across studies.

The simulations and *in-vivo* results suggest that the proposed analysis method could significantly increase the utility of dc-fMRI, reducing the number of participants needed to detect a group difference in oxygen metabolism or oxygen extraction fraction and offering more physiological interpretability of metabolic differences or alteration due to a stimulus. In addition, the significant reduction in processing time and the improved robustness of the individual parameter maps reduces two of the hurdles restricting clinical implementation of such techniques.

DATA AVAILABILITY STATEMENT

The python code for the machine learning implementation proposed in this article can be found in the fml_pMRI repository https://zenodo.org/badge/latestdoi/189416118. We do not have ethical consent to make the *in-vivo* datasets acquired for this study publically available.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by School of Psychology Research Ethics Committee Cardiff University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MG wrote the manuscript, developed and implemented the methods, and analyzed the *in-vivo* data. HC acquired and processed the data and edited the manuscript. TO created and provided the code used in the prototype pseudo-continuous arterial spin labeling pulse sequence. FF assisted in the implementation of the prototype arterial spin labeling pulse sequence. VT was the chief investigator overseeing the study design and data collection for a subset of healthy controls. KM and RW was the principal investigator overseeing the study design and data collection for a subset of healthy controls. All authors reviewed and edited the manuscript prior to submission.

FUNDING

This study was funded by Wellcome Strategic Award, Multi-scale and multi-modal assessment of coupling in the healthy and diseased brain, grant reference 104943/Z/14/Z (RW, MG, and HC). RW was also supported by the Higher Education Funding Council for Wales. KM was supported by Wellcome grant 200804/Z/16/Z. TO was supported by the Royal Academy of Engineering and Wellcome Centre for Integrative Neuroimaging is supported by core funding from Wellcome (203139/Z/16/Z).

ACKNOWLEDGMENTS

We would like to thank Wellcome for supporting this work: Wellcome Strategic Award, Multi-scale and multi-modal assessment of coupling in the healthy and diseased brain, grant reference 104943/Z/14/Z.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frai.2020.00012/full#supplementary-material

Supplementary Table 1 | Demographic data for healthy volunteers.

REFERENCES

Alsop, D. C., Detre, J. A., Golay, X., Gunther, M., Hendrikse, J., Hernandez-Garcia, L., et al. (2015). Recommended implementation of arterial spin-labeled perfusion MRI for clinical applications: a consensus of the ISMRM perfusion study group and the European consortium for ASL in dementia. *Magn. Reson. Med.* 73, 102–116. doi: 10.1002/mrm.25197

Bernier, J. L., Ortega, J., Ros, E., Rojas, I., and Prieto, A. (1999). A new measurement of noise immunity and generalization ability for MLPs. *Int. J. Neural. Syst.* 9, 511–521. doi: 10.1142/S0129065799000551

Blockley, N. P., Griffeth, V. E., Germuska, M. A., Bulte, D. P., and Buxton, R. B. (2013). An analysis of the use of hyperoxia for measuring venous cerebral blood volume: comparison of the existing method with a new analysis approach. *Neuroimage* 72, 33–40. doi: 10.1016/j.neuroimage.2013.01.039

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

Bulte, D. P., Kelly, M., Germuska, M., Xie, J., Chappell, M. A., Okell, T. W., et al. (2012). Quantitative measurement of cerebral physiology using respiratory-calibrated MRI. *Neuroimage* 60, 582–591. doi: 10.1016/j.neuroimage.2011.12.017

Chappell, M. A., MacIntosh, B. J., Donahue, M. J., Gunther, M., Jezzard, P., and Woolrich, M. W. (2010). Separation of macrovascular signal in multi-inversion time arterial spin labelling MRI. *Magn. Reson. Med.* 63, 1357–1365. doi: 10.1002/mrm.22320

Coles, J. P., Fryer, T. D., Bradley, P. G., Nortje, J., Smielewski, P., Rice, K., et al. (2006). Intersubject variability and reproducibility of 15O PET studies. *J. Cereb. Blood Flow. Metab.* 26, 48–57. doi: 10.1038/sj.jcbfm.9600179

De Vis, J. B., Petersen, E. T., Bhogal, A., Hartkamp, N. S., Klijn, C. J., Kappelle, L. J., et al. (2015). Calibrated MRI to evaluate cerebral hemodynamics in patients with an internal carotid artery occlusion. *J. Cereb. Blood Flow. Metab.* 35, 1015–1023. doi: 10.1038/jcbfm.2015.14

Frackowiak, R. S., Herold, S., Petty, R. K., and Morgan-Hughes, J. A. (1988). The cerebral metabolism of glucose and oxygen measured with positron tomography in patients with mitochondrial diseases. *Brain* 111, 1009–1024. doi: 10.1093/brain/111.5.1009

Frackowiak, R. S., Lenzi, G. L., Jones, T., and Heather, J. D. (1980). Quantitative measurement of regional cerebral blood flow and oxygen metabolism in man using 15O and positron emission tomography: theory, procedure, and normal values. *J. Comput. Assist. Tomogr.* 4, 727–736. doi: 10.1097/00004728-198012000-00001

Frau-Pascual, A., Vincent, T., Sloboda, J., Ciuciu, P., and Forbes, F. (2014). "Physiologically informed bayesian analysis of ASL fMRI data," in *Bayesian and Graphical Models for Biomedical Imaging. Lecture Notes in Computer Science*, Vol. 8677 (Cham: Springer), 37–48. doi: 10.1007/978-3-319-12289-2_4

Friston, K. J., Holmes, A. P., Poline, J. B., Grasby, P. J., Williams, S. C. R., Frackowiak, R. S. J., et al. (1995). Analysis of fmri time-series revisited. *Neuroimage* 2, 45–53. doi: 10.1006/nimg.1995.1007

Gauthier, C. J., Desjardins-Crepeau, L., Madjar, C., Bherer, L., and Hoge, R. D. (2012). Absolute quantification of resting oxygen metabolism and metabolic reactivity during functional activation using QUO2 MRI. *Neuroimage* 63, 1353–1363. doi: 10.1016/j.neuroimage.2012.07.065

Gauthier, C. J., and Hoge, R. D. (2013). A generalized procedure for calibrated MRI incorporating hyperoxia and hypercapnia. *Hum. Brain Mapp.* 34, 1053–1069. doi: 10.1002/hbm.21495

Germuska, M., Chandler, H. L., Stickland, R. C., Foster, C., Fasano, F., Okell, T. W., et al. (2019). Dual-calibrated fMRI measurement of absolute cerebral metabolic rate of oxygen consumption and effective oxygen diffusivity. *Neuroimage* 184, 717–728. doi: 10.1016/j.neuroimage.2018.09.035

Germuska, M., Merola, A., Murphy, K., Babic, A., Richmond, L., Khot, S., et al. (2016). A forward modelling approach for the estimation of oxygen extraction fraction by calibrated fMRI. *Neuroimage* 139, 313–323. doi: 10.1016/j.neuroimage.2016.06.004

Germuska, M., and Wise, R. G. (2019). Calibrated fMRI for mapping absolute $CMRO_2$: practicalities and prospects. *Neuroimage* 187, 145–153. doi: 10.1016/j.neuroimage.2018.03.068

Geurts, P., Ernst, D., and Wehenkel, L. (2006). Extremely randomized trees. *Mach. Learn.* 63, 3–42. doi: 10.1007/s10994-006-6226-1

Gjedde, A. (2002). Cerebral blood flow change in arterial hypoxemia is consistent with negligible oxygen tension in brain mitochondria. *Neuroimage* 17, 1876–1881. doi: 10.1006/nimg.2002.1272

Hayashi, T., Watabe, H., Kudomi, N., Kim, K. M., Enmi, J., Hayashida, K., et al. (2003). A theoretical model of oxygen delivery and metabolism for physiologic interpretation of quantitative cerebral blood flow and metabolic rate of oxygen. *J. Cereb. Blood Flow. Metab.* 23, 1314–1323. doi: 10.1097/01.WCB.0000090506.76664.00

Hertel, L., Phan, H., and Mertins, A. (2016). "Comparing time and frequency domain for audio event recognition using deep learning," in *2016 International Joint Conference on Neural Networks (IJCNN)* (Vancouver, BC), 3407–3411. doi: 10.1109/IJCNN.2016.7727635

Hoge, R. D. (2012). Calibrated fMRI. *Neuroimage* 62, 930–937. doi: 10.1016/j.neuroimage.2012.02.022

Hyder, F., Herman, P., Bailey, C. J., Moller, A., Globinsky, R., Fulbright, R. K., et al. (2016). Uniform distributions of glucose oxidation and oxygen extraction in gray matter of normal human brain: no evidence of regional differences of aerobic glycolysis. *J. Cereb. Blood Flow. Metab.* 36, 903–916. doi: 10.1177/0271678X15625349

Ibaraki, M., Shinohara, Y., Nakamura, K., Miura, S., Kinoshita, F., and Kinoshita, T. (2010). Interindividual variations of cerebral blood flow, oxygen delivery, and metabolism in relation to hemoglobin concentration measured by positron emission tomography in humans. *J. Cereb. Blood Flow. Metab.* 30, 1296–1305. doi: 10.1038/jcbfm.2010.13

Ishii, K., Kitagaki, H., Kono, M., and Mori, E. (1996). Decreased medial temporal oxygen metabolism in Alzheimer's disease shown by PET. *J. Nucl. Med.* 37, 1159–1165.

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., and Smith, S. M. (2012). Fsl. *Neuroimage* 62, 782–790. doi: 10.1016/j.neuroimage.2011.09.015

Lajoie, I., Nugent, S., Debacker, C., Dyson, K., Tancredi, F. B., Badhwar, A., et al. (2017). Application of calibrated fMRI in Alzheimer's disease. *Neuroimage Clin.* 15, 348–358. doi: 10.1016/j.nicl.2017.05.009

Lebrun-Grandie, P., Baron, J. C., Soussaline, F., Loch'h, C., Sastre, J., and Bousser, M. G. (1983). Coupling between regional blood flow and oxygen utilization in the normal human brain. A study with positron tomography and oxygen 15. *Arch. Neurol.* 40, 230–236. doi: 10.1001/archneur.1983.04050040060010

Leenders, K. L., Perani, D., Lammertsma, A. A., Heather, J. D., Buckingham, P., Healy, M. J., et al. (1990). Cerebral blood flow, blood volume and oxygen utilization. Normal values and effect of age. *Brain* 113, 27–47. doi: 10.1093/brain/113.1.27

Merola, A., Germuska, M. A., Murphy, K., and Wise, R. G. (2018). Assessing the repeatability of absolute CMRO2, OEF and haemodynamic measurements from calibrated fMRI. *Neuroimage* 173, 113–126. doi: 10.1016/j.neuroimage.2018.02.020

Merola, A., Germuska, M. A., Warnert, E. A., Richmond, L., Helme, D., Khot, S., et al. (2017). Mapping the pharmacological modulation of brain oxygen metabolism: the effects of caffeine on absolute $CMRO_2$ measured using dual calibrated fMRI. *Neuroimage* 155, 331–343. doi: 10.1016/j.neuroimage.2017.03.028

Merola, A., Murphy, K., Stone, A. J., Germuska, M. A., Griffeth, V. E. M., Blockley, N. P., et al. (2016). Measurement of oxygen extraction fraction (OEF): an optimized BOLD signal model for use with hypercapnic and hyperoxic calibration. *Neuroimage* 129, 159–174. doi: 10.1016/j.neuroimage.2016.01.021

Mesejo, P., Saillet, S., David, O., Bénar, C., Warnking, J. M., and Forbes, F. (2015). "Estimating biophysical parameters from BOLD signals through evolutionary-based optimization," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Lecture Notes in Computer Science,* eds N. Navab, J. Hornegger, W. Wells, and A. Frangi (Cham: Springer), 9350, 528–535 doi: 10.1007/978-3-319-24571-3_63

Miles, K. A., and Williams, R. E. (2008). Warburg revisited: imaging tumour blood flow and metabolism. *Cancer Imaging* 8, 81–86. doi: 10.1102/1470-7330.2008.0011

Mintun, M. A., Lundstrom, B. N., Snyder, A. Z., Vlassenko, A. G., Shulman, G. L., and Raichle, M. E. (2001). Blood flow and oxygen delivery to human brain during functional activity: theoretical modeling and experimental data. *Proc. Natl. Acad. Sci. U.S.A.* 98, 6859–6864. doi: 10.1073/pnas.111164398

Morris, E. A., Juttukonda, M. R., Lee, C. A., Patel, N. J., Pruthi, S., Donahue, M. J., et al. (2018). Elevated brain oxygen extraction fraction in preterm newborns with anemia measured using noninvasive MRI. *J. Perinatol.* 38, 1636–1643. doi: 10.1038/s41372-018-0229-1

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830. doi: 10.5555/1953048.2078195

Powers, W. J., Videen, T. O., Markham, J., Walter, V., and Perlmutter, J. S. (2011). Metabolic control of resting hemispheric cerebral blood flow is oxidative, not glycolytic. *J. Cereb. Blood Flow. Metab.* 31, 1223–1228. doi: 10.1038/jcbfm.2011.5

Reeve, H. W. J., and Brown, G. (2018). Diversity and degrees of freedom in regression ensembles. *Neurocomputing* 298, 55–68. doi: 10.1016/j.neucom.2017.12.066

Rosenfeld, A., and Kak, A. C. (1982). *Digital Picture Processing*. Vol. I & II. New York, NY: Academic Press.

Ruijters, D., and Thevenaz, P. (2012). GPU Prefilter for accurate Cubic B-spline interpolation. *Comput. J.* 55, 15–20. doi: 10.1093/comjnl/bxq086

Safar, P. (1988). Resuscitation from clinical death: pathophysiologic limits and therapeutic potentials. *Crit Care Med.* 16, 923–941. doi: 10.1097/00003246-198810000-00003

Scheinberg, P., and Stead, E. A. (1949). The cerebral blood flow in male subjects as measured by the nitrous oxide technique. Normal values for blood flow, oxygen utilization, glucose utilization, and peripheral resistance, with observations on the effect of tilting and anxiety. *J. Clin. Invest.* 28, 1163–1171. doi: 10.1172/JCI102150

Sollich, P., and Krogh, A. (1996). Learning with ensembles: how over-fitting can be useful. *Adv. Neural Inform. Process. Syst.* 8, 190–196.

Vafaee, M. S., and Gjedde, A. (2000). Model of blood-brain transfer of oxygen explains nonlinear flow-metabolism coupling during stimulation of visual cortex. *J. Cereb. Blood Flow. Metab.* 20, 747–754. doi: 10.1097/00004647-200004000-00012

Verweij, B. H., Amelink, G. J., and Muizelaar, J. P. (2007). Current concepts of cerebral oxygen transport and energy metabolism after severe traumatic brain injury. *Prog. Brain. Res.* 161, 111–124. doi: 10.1016/S0079-6123(06)61008-X

Wise, R. G., Harris, A. D., Stone, A. J., and Murphy, K. (2013). Measurement of OEF and absolute CMRO2: MRI-based methods using interleaved and combined hypercapnia and hyperoxia. *Neuroimage* 83, 135–147. doi: 10.1016/j.neuroimage.2013.06.008

Yue, J. B., Feng, H. K., Yang, G. J., and Li, Z. H. (2018). A comparison of regression techniques for estimation of above-ground winter wheat biomass using near-surface spectroscopy. *Rem. Sens.* 10:66. doi: 10.3390/rs10010066

Conflict of Interest: FF was employed by company Siemens Healthcare Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Artificial Intelligence

CORRECTION
published: 12 July 2021
doi: 10.3389/frai.2021.614245

Corrigendum: A Frequency-Domain Machine Learning Method for Dual-Calibrated fMRI Mapping of Oxygen Extraction Fraction (OEF) and Cerebral Metabolic Rate of Oxygen Consumption ($CMRO_2$)

Michael Germuska[1], Hannah Louise Chandler[1], Thomas Okell[2], Fabrizio Fasano[3], Valentina Tomassini[1,4,5], Kevin Murphy[1] and Richard G. Wise[1,5,6]*

[1]Cardiff University Brain Research Imaging Centre (CUBRIC), Department of Psychology, Cardiff University, Cardiff, United Kingdom, [2]FMRIB, Nuffield Department of Clinical Neurosciences, Wellcome Centre for Integrative Neuroimaging, University of Oxford, Oxford, United Kingdom, [3]Siemens Healthcare Ltd., Camberley, United Kingdom, [4]Division of Psychological Medicine and Clinical Neurosciences, Cardiff University School of Medicine, Cardiff, United Kingdom, [5]Department of Neuroscience, Imaging and Clinical Sciences, "G. D'Annunzio University" of Chieti-Pescara, Chieti, Italy, [6]Institute for Advanced Biomedical Technologies, "G. D'Annunzio University" of Chieti-Pescara, Chieti, Italy

OPEN ACCESS

Edited by:
Shuihua Wang,
University of Leicester,
United Kingdom

Reviewed by:
Daniel Bulte,
University of Oxford, United Kingdom

***Correspondence:**
Michael Germuska
germuskam@cardiff.ac.uk

Specialty section:
This article was submitted to Medicine and Public Health, a section of the journal Frontiers in Artificial Intelligence

Received: *05 October 2020*
Accepted: *29 June 2021*
Published: *12 July 2021*

Citation:
Germuska M, Chandler HL, Okell T, Fasano F, Tomassini V, Murphy K and Wise RG (2021) Corrigendum: A Frequency-Domain Machine Learning Method for Dual-Calibrated fMRI Mapping of Oxygen Extraction Fraction (OEF) and Cerebral Metabolic Rate of Oxygen Consumption ($CMRO_2$). Front. Artif. Intell. 4:614245. doi: 10.3389/frai.2021.614245

Keywords: calibrated-fMRI, oxygen extraction fraction, $CMRO_2$, OEF, machine learning

A Corrigendum on

A Frequency-Domain Machine Learning Method for Dual-Calibrated fMRI Mapping of Oxygen Extraction Fraction (OEF) and Cerebral Metabolic Rate of Oxygen Consumption ($CMRO_2$)

by Germuska M., Chandler H., Okell T., Fasano F., Tomassini V., Murphy K., Wise R. G. (2020). Front. Artif. Intell. 3:12. doi: 10.3389/frai.2020.00012

In the original article, there was a mistake in **Table 3** as published. The labels for [Hb] and OEF_0 were swapped. The corrected **Table 3** appears below.

In the original article, there was an error. Reference to **Table 3** in the text had the labels for [Hb] and OEF_0 swapped. A correction has been made to **Results**, ***In-vivo***, paragraph 5:

"**Table 3** reports the results of a bivariate regression of OEF against [Hb] and CBF for both analysis methods. The slopes of the relationship between OEF and [Hb] are similar to that reported in healthy subjects by Ibaraki et al. (2010), −1.75 Hb (g/dL). As per Ibaraki et al. the relationship between CBF and OEF did not reach significance ($p = 0.44$) for the ML approach, however a significant negative correlation was observed in the rNLS analysis ($p = 0.005$). A univariate analysis of CMRO2,0 against CBF0 is consistent with that observed in healthy controls by Powers et al. (2011) ($\beta1 = 0.2$) for both

TABLE 3 | Results of a bivariate regression of OEF_0 against CBF_0 and [Hb] for 30 healthy volunteers analyzed with the ML (ensemble of MLPs) and rNLS fitting methods.

Predictor	ML β1 (p value)	rNLS β1 (p value)
[Hb]	−1.42 (0.001)	−2.23 (0.001)
CBF	−0.07 (0.44)	−0.37 (0.005)
Intercept	61.95 (<0.001)	89.48 (<0.001)

analysis methods, $\beta1 = 0.32$ ($p < 0.001$) and $\beta1 = 0.24$ ($p < 0.001$) for the ML and rNLS approaches respectively."

The authors apologize for these errors and state that this does not change the scientific conclusions of the article in any way. The original article has been updated.

frontiers in Public Health

REVIEW
published: 12 May 2020
doi: 10.3389/fpubh.2020.00164

Generative Adversarial Networks and Its Applications in Biomedical Informatics

Lan Lan[1†], Lei You[2†], Zeyang Zhang[3†], Zhiwei Fan[4†], Weiling Zhao[2], Nianyin Zeng[5*], Yidong Chen[6] and Xiaobo Zhou[2]*

[1] West China Biomedical Big Data Center, West China Hospital, Sichuan University, Chengdu, China, [2] Center for Computational Systems Medicine, School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston, TX, United States, [3] Department of Computer Science and Technology, College of Electronics and Information Engineering, Tongji University, Shanghai, China, [4] Department of Epidemiology and Health Statistics, West China School of Public Health and West China Fourth Hospital, Sichuan University, Chengdu, China, [5] Department of Instrumental and Electrical Engineering, Xiamen University, Fujian, China, [6] Department of Computer Science and Technology, College of Computer Science, Sichuan University, Chengdu, China

OPEN ACCESS

Edited by:
Shuihua Wang,
University of Leicester,
United Kingdom

Reviewed by:
Robertas Damasevicius,
Kaunas University of Technology, Lithuania
Fuhai Li,
Washington University in St. Louis, United States

****Correspondence:***
Lan Lan
lanl@scu.edu.cn
Nianyin Zeng
zny@xmu.edu.cn

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *30 December 2019*
Accepted: *17 April 2020*
Published: *12 May 2020*

Citation:
Lan L, You L, Zhang Z, Fan Z, Zhao W, Zeng N, Chen Y and Zhou X (2020) Generative Adversarial Networks and Its Applications in Biomedical Informatics. Front. Public Health 8:164. doi: 10.3389/fpubh.2020.00164

The basic Generative Adversarial Networks (GAN) model is composed of the input vector, generator, and discriminator. Among them, the generator and discriminator are implicit function expressions, usually implemented by deep neural networks. GAN can learn the generative model of any data distribution through adversarial methods with excellent performance. It has been widely applied to different areas since it was proposed in 2014. In this review, we introduced the origin, specific working principle, and development history of GAN, various applications of GAN in digital image processing, Cycle-GAN, and its application in medical imaging analysis, as well as the latest applications of GAN in medical informatics and bioinformatics.

Keywords: Generative Adversarial Networks (GAN), generator, discriminator, data augmentation, image conversion, biomedical applications

INTRODUCTION

Generative Adversarial Networks (GAN) was introduced into the field of deep learning by Goodfellow et al. (1). As can be seen from its name, GAN, a form of generative models, is trained in an adversarial setting deep neural network. More specifically, GAN learns the generative model of data distribution through adversarial methods. GAN is the most successful generative model developed in recent years and has become one of the hottest research directions in the field of artificial intelligence. Because of its excellent performance, GAN attracts great attention since it was proposed. It is especially important that GAN can not only be used as a generative model with excellent performance, but also its inspiring adversarial learning idea penetrates deeply into all aspects of deep learning, resulting in a series of new research directions and various applications (2).

The basic function of GAN is to train a generator and discriminator in an adversarial way. Based on different requirements of projects, either a stronger generator or a more sensitive discriminator is designed as the target goal. In this manuscript, we focus on the generation purpose of GAN used in four areas: digital image processing, medical image processing, medical informatics, and its latest applications in omic data. The generation purpose can be further categorized into data simulation (3), data augmentation for small dataset (4), style transformation (5), and gene data simulation (6). The great successful applications of GAN in medical image generation (7, 8) and cell gene imputation (6) motivated us to review the literatures in these four sub areas, rather than

just focusing on the digital image processing field. We searched in the top conferences of computer science and Google Scholar with keywords related to GAN. Through screening the literature abstracts by our team in digital image processing, medical imaging analysis, medical informatics, and bioinformatics, respectively, the literature that was very relevant to our subject was retained for full text reading. The contents of these eligible literatures are summarized below.

A BRIEF OVERVIEW OF GAN

Origin of GAN

In general, deep learning models can be divided into discriminant models and generative models (9). In the perspective of the probability and statistical theory, a discriminant model is a method of modeling the relationship between unknown data y and known data x. A generating model refers to a model that can randomly generate observations, especially under the condition of given some implicit parameters (10). Due to the invention of algorithms such as Back Propagation (BP) and Dropout, the discriminant model has been evolved rapidly. The development of the generative model is lagged due to the difficulty of modeling, though the generative model has a pivotal role in the history of machine learning. When processing large amounts of data, such as images, speech, text, genomics, etc., the generative models can help us simulate the distribution of these high-dimensional data. It will be beneficial for many applications, such as super-resolution, data augmentation, image and medical image conversion, caption generation, electronic health records data generation, biomedical data generation, data imputation, and other ill-posed problems (11–15).

Likelihood describes the probability of the event under different conditions when the results are known (16). Sometimes we may not know the distribution function, but we know the observed data. Therefore, the maximum likelihood estimation is applied to evaluate model parameters using the observed data. Traditional generative models such as Restricted Boltzmann Machine (RBM) (17, 18), Gaussian Mixture Model (GMM) (19), Naive Bayes Model (NBM) (20), Hidden Markov Model (HMM) (20) and so on, are mostly based on maximum likelihood estimate. However, while the explicitly defined probability density function brings computational tractability, maximum likelihood estimation may not represent the complexity of the actual data distribution and cannot learn the high-dimensional data distributions. The majority of generative models require the utilization of Markov chains. GAN uses latent codes to express latent dimensions, control data implicit relationships, etc. and does not require Markov chains (21). Adversarial networks can represent very sharp, even degenerate distributions, while Markov chain-based approaches require somewhat ambiguous distributions so that the chains can be mixed between patterns. Various types of loss functions can be integrated into GAN models. This allows different types of loss functions to be designed for different tasks, all of which can be learned and optimized under the GAN framework. GAN is also a nonparametric modeling method and does not require an approximate distribution of training data to be defined in advance. When probability density is not computable, some traditional generative models that rely on the statistical interpretation of data cannot be used for learning and application. But GAN can still be used in such cases.

Specific Principles of GAN

In this section, we will introduce the architecture and specific principles of GAN. Basic GAN model is composed of an input vector, a generator, and a discriminator. The generator and discriminator are implicit function expressions, usually implemented by deep neural networks (22).

We use abstract mathematical language to explain the basic principles of the GAN. The fixed distribution $P_{data}(x)$ is usually calculated based on the assumption that the data distribution for the training sample x is P_{data}. However, this distribution is difficult to be determined. The traditional methods assume that the distribution $P_{data}(x)$ obeys a Gaussian mixture distribution and uses the maximum likelihood as the solution. However, when the model is complicated, it is often unable to calculate and the resulting performance is limited (23). This is due to the limited expression ability of the Gaussian distribution itself. Thus, neural networks were proposed to define the distribution $P_g(x)$. The generator is a neural network with parameter θ^g. It collects the random variable z from the prior distribution and maps it to the pseudo-sample distribution through the neural network, that is, the generated data is recorded as $G(z)$ and the data distribution is recorded as $P_g(z)$. The input z usually uses Gaussian noise, which is a random variable or a random variable in the potential space. According to θ^g, a simple input distribution can be used to generate various complex distributions. The $P_g(x)$ generated by the generator and the real image distribution $P_{data}(x)$ should be as similar as possible (24). So, for the generator, the target is to find a G^* as shown below.

$$G^* = arg \min_G Div(P_g, P_{data}) \tag{1}$$

Then the next question is how to calculate the difference between the two distributions. If the form of $P_{data}(x)$ and $P_g(x)$ is known, it can be calculated to make $P_{data}(x)$ and $P_g(x)$ get close. Although we don't know the specific distribution, we can sample from it. So, GAN proposed a very magical way, discriminator, to calculate the difference between the two distributions. The discriminator was defined by the original GAN as a binary classifier (25) with θ^d. During training, when the input is a real sample x, the output of discriminator should be 1, otherwise, the output goes to 0. For defining discriminator, Goodfellow et al. (1) used binary cross entropy function, which is commonly used for binary classification problems.

$$\text{Loss} = -(y \log(\hat{y}) + (1-y)\log(1-\hat{y})) \tag{2}$$

Where $\hat{y}$ is the probability that the model prediction sample is a positive example, and y is the sample label. If the sample belongs to a positive example, the value is 1; otherwise, the value is 0. A specific sample may come either from the real distribution or the generated distribution. The positive and negative cases

are substituted into P_{data} and P_g, respectively. The whole object function for discriminator is:

$$\mathrm{V}(\mathrm{G},\mathrm{D}) = E_{x\sim P_{data}}\left[\log D(x)\right] + E_{x\sim P_g}[\log(1-D(x))] \quad (3)$$

By merging Equation (1) into (3), the objective function of the basic GAN is defined by Equation (4):

$$\min_G \max_D V(G,D) = \min_G \max_D E_{x\sim P_{data}}\left[\log D(x)\right] + E_{z\sim P_z}[\log(1-D(G(z)))] \quad (4)$$

By optimizing this objective function, we can get a GAN model. GAN's training can be regarded as a min–max optimization process. The generator wants to deceive the discriminator, so it tries to maximize discriminator's output when a fake sample is presented to the discriminator. Instead, the discriminator attempts to distinguish the difference between real and false samples. Consequently, discriminator tries to maximize $V(G, D)$ while generator tries to minimize $V(G, D)$, thus forming the minimax relationship. During the training of GAN, the parameters of G (θ^g) and D (θ^d) are continuously updated. When the generator is undergoing training, the parameters of the discriminator are fixed. The data generated by the generator is marked as fake and input into the discriminator. The error is calculated between the output of the discriminator $D(G(z))$ and the sample label, and the parameters of generator are updated using the error of BP algorithm. When the discriminator is undergoing training, the parameters of the generator being fixed. Discriminator gets positive sample x from the real data set, and the generator generates a negative sample $G(z)$. The output of the discriminator and sample labels are used to calculate the error. Finally, the parameters of the discriminator are updated by the error of BP algorithm.

Ideally, the generator and discriminator are in equilibrium when $P_{data}(x) = P_g(x)$. When the generator is fixed, we can take the derivative of $V(D, G)$ to find the optimal discriminator $D^*(x)$, as shown in the Equation (5).

$$D^*(x) = \frac{P_g(x)}{P_g(x) + P_{data}(x)} \quad (5)$$

By substituting the optimal discriminator in the Equation (3).

$$\max_D V(G,D) = -2\log 2 + 2JSD(P_{data}(x) \parallel P_g(x)) \quad (6)$$

The objective function can be further calculated as optimizing the JS divergence of $P_{data}(x)$ and $P_g(x)$ under the optimal discriminator (26).

Development History of GAN

GAN is an excellent generative model. However, the original GAN model has many problems, such as the vanishing gradient, difficulty in training, and poor diversity (27). Many efforts have been devoted to obtaining better GANs through different optimization methods. Therefore, since 2014, theories and articles related to GAN have emerged in an endless stream, and many new GANs-based models have been proposed to improve the stability and quality of the generated results (28).

A number of review articles have summarized and classified the current GAN-related models (22, 24, 29). Creswell et al. (22) classified the evolution of GAN models from the aspects of architectural development and loss function improvement. Hong et al. (29) summarized the development of GAN models from the aspects of theoretical analysis, supervised, unsupervised, and common problems. Guo et al. (24) focused on the improvement of the model structure, the expansion of the theory, the novel application and so on. We will introduce several common improvements of GAN here.

Conditional Generative Adversarial Networks (CGAN)

CGAN is an improved GAN model proposed by Mirza et al. (30). Unlike the original GAN, CGAN uses a supervised approach increasing controllability of generated results. CGAN takes the random noise z and the category label c as inputs of the generator and the generated sample/real sample and category label as inputs of the discriminator to learn the correlation between labels and images. By introducing a conditional variable y into the modeling and adding conditions to the model with additional information y, the data generation process can be guided.

Deep Convolutional Generative Adversarial Networks (DCGAN)

One year after the first GAN paper was published, researchers found that the GAN model was unstable and required a lot of training skills. In 2015, Radford et al. (31) proposed an upgraded version of the GAN architecture, named DCGAN. The authors of DCGAN improved the architecture of the original GAN with deep convolutional networks (CNNs). So far, DCGAN's network structure is still widely used and is the hottest GAN architecture and a milestone in the history of GAN. Compared with the original GAN, DCGAN almost completely uses the convolution layer instead of the fully connected layer. The discriminator is almost symmetric with the generator. The entire network does not have pooling layers and up-sampling layers. DCGAN also used Batch Normalization algorithm to solve the problem of vanishing gradient.

f-GAN

The objective function of the original GAN can be seen as minimizing the JS divergence between two distributions. In fact, there are many ways to measure the distance between two distributions, and JS divergence is just one of them. Defining different distance metrics can result in different objective functions. Nowozin et al. (32) applied f-divergence to GAN (f-GAN) for training generative neural samplers. The f-divergence is a function $D_f(P \parallel Q)$ that measures the difference between two probability distributions P and Q. Under the framework of f-divergence, f-GAN generalizes various divergences so that the corresponding GAN target can be derived for a specific divergence. Many common divergences (33), such as KL-divergence, Hellinger distance, and total variation distance, are the special cases of f-divergence, coinciding with a particular choice of f. Many improvements in GAN training

TABLE 1 | Literatures for the application of GAN in image processing.

References	Model	Public dataset
Mirza and Osindero (30)	CGAN	MNIST
Radford et al. (31)	DCGAN	LSUN, IMAGENET-1K
Nowozin et al. (32)	*f*-GAN	MNIST Digits, LSUN
Zhao et al. (34)	EBGAN	MNIST digit, LSUN, CelebA, ImageNet
Arjovsky et al. (26)	WGAN	LSUN-Bedrooms
Karras et al. (37)	proGAN	CelebA, LSUN
Ledig et al. (12)	SRGAN	Set5, Set14, BSD100
Pathak et al. (38)	Context encoder	Paris Street View, ImageNet

stability are achieved by using different distance metrics between distributions, such as Energy-based GAN (EBGAN) (34), Least Squares GAN (LSGAN) (35), etc.

Wasserstein Generative Adversarial Networks (WGAN)

WGAN mainly improved GAN from the perspective of the loss function. WGAN theoretically explained the reason for the instability of GAN training, that is, cross entropy (JS divergence) is not suitable for measuring the distance between distributions with disjoint parts. Therefore, WGAN proposed a new distance measurement method, Earth Moving Distance, also known as Wasserstein distance or optimal transmission distance, which refers to the minimum transmission quality that converts the probability distribution q to p (probability density is called probability quality in discrete cases) (26, 36). The superiority of Wasserstein distance compared to KL divergence and JS divergence is that even if two distributions do not overlap, Wasserstein distance can still reflect their distance. The theoretical derivation and interpretation of WGAN are quite complicated. The authors of WGAN (26) pointed out that the use of Wasserstein distances needs to satisfy a strong continuity condition, i.e. Lipchitz continuity.

In short, GAN still has many unresolved problems and can be further improved in various aspects.

APPLICATION OF GAN IN IMAGE PROCESSING

GAN is widely used in virtual image generation (**Table 1**). Whether it is a face image, a room scene image, a real image (37) such as a flower or an animal, or an artistic creation image such as an anime character (39), it can be learned using GAN to generate new similar images (**Figure 1**). GAN is fully utilizing its unique advantages and has evolved from the original GAN to the progressively growing GAN (proGAN). Its imaging generation capability has been greatly improved from 32 × 32 resolution to 2K true and false HD resolution (37).

Image super-resolution task (SR) is to generate high-resolution images from low-resolution images. Image super-resolution algorithms are important in areas such as video surveillance, medical diagnostics, and remote sensing applications. Super-resolution problem is actually an ill-posed problem because the lost high-frequency details are difficult to recover during the resolution of the image. Traditional methods are generally interpolated, and interpolation inevitably creates blurring. However, GAN can learn the distribution of high-resolution images to a certain extent, so that high-quality images with better quality can be generated. CNN has also achieved very good results in single-frame super-resolution reconstruction and can achieve a higher peak signal-to-noise ratio (PSNR) (40, 41). However, most of them use MSE as the objective function. Although a higher peak signal-to-noise ratio can be achieved using MSE, when the image down-sampling is higher (four times), the reconstructed picture will be too smooth and lose details. Thus, in 2016, Super Resolution GAN (SRGAN) was proposed by Ledig and others of Twitter. SRGAN was the first to propose the application of GAN to super-resolution reconstruction (12). The generated model of SRGAN takes a blurred low-resolution image as input and outputs a clear image with high resolution. The discriminant model of SRGAN determines whether the input image is a "true high-resolution image" or a high-resolution image converted from a low-resolution image. This greatly simplifies the learning process of the image super-resolution model. Because traditionally conducting an image super-resolution needs to model some high-frequency details, but here the purpose of generating model training is simplified to the confusing discriminant model. Compared to previous results based on deep learning models for image super-resolution such as SRResNet, etc., GAN's images can provide more details.

Image inpainting refers to the process of reconstructing missing or damaged parts of images and videos. It involves in image editing and image generation and is a process of artificially filling a region where information on the image is missing according to certain rules. Conventional image inpainting methods typically utilize undamaged image information to estimate missing portions and autofill the missing parts. Therefore, a structure-based partial differential equation (PDE) image restoration algorithms were proposed. The repair process of these models is similar to the diffusion phenomenon of physics, and the key and the difficulty lies in how to build a diffusion model. When the image damage area is large, the repair effect plummets. Texture-based image inpainting algorithms emerged as the times require, and the effect based on partial differential equations was improved to some extent. Because the traditional image inpainting algorithms depend on the structure or texture information of the image, they cannot meet basic repair requirements, and often cannot achieve satisfactory results, when the image semantic information is missing. Deep learning has strong learning ability and can learn advanced features from images, so the inpainting problem can be solved with such features (42–44). Image inpainting is a problem between image editing and image generation, so using the GAN model can solve this problem well. The solution for using GAN is to input an image with missing part to the generator. The generator will use this missing input image to generate a new complete image. The discriminator will learn to judge whether

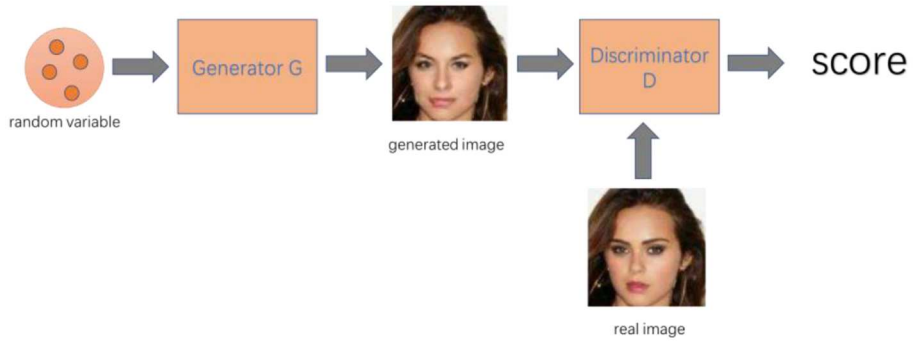

FIGURE 1 | Digital image diagram of GAN.

this image is realistic enough and feedback to the generator. Through continuous training to optimize, the generator can finally generate a complete image that is sufficiently realistic. Then the inpainting was finished. Context encoder (38) is a pioneering work of deep learning in the field of image inpainting. Pathak et al. (38) trains an encoder-decoder and combines adversarial network loss to predict the missing portion according to the context pixel and structural semantics of the missing area. The network is able to obtain a reasonable image structure and can quickly and accurately evaluate the repair results (45–47). Because GAN's generator and discriminator can be any form of neural network, different network architectures can be selected for solving different problems. The autoencoder model was used as the generator part in our previous work (3). This is because the autoencoder is also an important generative model. It encodes the input data and generates new data through the decoding operation. It can retain the characteristics of the original input and introduce the newly generated part. Therefore, it can not only keep the undefected part of the inpainting input, but also generate some data for filling in the missing area by using autoencoder. Our discriminator is a simple binary classification discriminant neural network whose input are the generated repaired image and the complete image in the original dataset. By learning to distinguish between the two, the generator is prompted to generate results that are more in line with the dataset, thereby completing the inpainting.

APPLICATION OF GAN IN MEDICAL IMAGE PROCESSING

Multiple factors such as time cost, labor cost, economic cost, etc. make it more difficult for researchers to acquire labeled medical images than normal images. However, there is a great demand for medical images by scholars nowadays. For example, the deep learning-based model can achieve a better performance in the fields of medical image segmentation, classification, registration, etc. than the hand-crafted features when dealing with a large amount of data (48). Traditional image augmentation methods can be used for its purpose. However, the generated images by traditional augmentation methods share a similar distribution with the original ones (49, 50). Those methods are not suitable for the need of generating more incidences among different patients. Accordingly, GAN is used more popular in medical image analysis, such as data augmentation and multi-modality image translations.

Recently, with the development of deep learning algorithms and the growing of labeled image datasets, convolutional neural networks (CNN)-based models (51) have achieved great success in many computer vision tasks, such as object detection (52), semantic segmentation (53), human action recognition (54) and so on.

Since 2014, many CNN-based medical image analysis works (55, 56) have shown great learning possibility when enough images are available for model training. The database like TCGA (57) supplies a large number of images for some common diseases. Since image acquisition and annotation is a time-consuming process, image data for many diseases remains scarce.

There are many deep learning models that are pre-trained on larger image datasets such as ImageNet (58), COCO (59), and so on. Transfer learning (60) uses limited labeled data for supervised training. In the transfer learning tasks, most of the weights of the model keep fixed and only the weights of the last several layers are fine-tuned on the new dataset. In this way, a well pre-trained deep neural network is applied in the medical image analysis.

Transfer learning may still suffer from lack of training images (61). As mentioned above, traditional data augmentation methods can only generate data that share a close distribution with the original ones. If the data set is too small, these methods almost have no effort on the data augmentation. GAN (26) supplies a solution to the lack of data in medical image analysis. In the following section, we will discuss the applications of single GAN and Cycle-GAN in medical image analysis (**Table 2**).

GAN Used in Medical Image Analysis

Pandey et al. (64) proposed a two-stage strategy to generate nuclei cell images and their corresponding masks based on GAN. In their first stage, a generator is trained to generate the synthesis masks from noise like conventional GAN model does. On their second stage, a conditional GAN utilized real mask and random noises to train a generator for synthesizing images. Finally, they

TABLE 2 | Literatures for the application of GAN in medical image processing.

References	Model	Public Dataset
Zhang et al. (62)	PAC-GAN	VIPeR, CUHK03, Market-1501
Dirvanauskas et al. (63)	GAN/medical	Miri TL
Pandey et al. (64)	Two-stage GAN/Medical	Kaggle Data Science bowl's first stage of competition
Frid-Adar et al. (65)	GAN	Private
Chen et al. (66, 67)	Dense GAN	A large publicly accessible brain structural MRI database
Mahapatra and Bozorgtabar (68)	Skip-connection GAN	http://www.eyepacs.com/
Yi and Babyn (69)	SAGAN	National Cancer Imaging Archive
Shitrit and Raviv (70)	GAN	Private
Zhu et al. (5)	Cycle GAN	Cityscapes, Google Maps, CMP Facade Database, UT Zappos50K, ImageNet
Wolterink et al. (71)	Cycle GAN	Private
Hiasa et al. (4)	Cycle GAN	Private
Huo et al. (72)	Cycle GAN	Private
Tanner et al. (73)	Cycle GAN	Private
Zhang et al. (74)	Cycle GAN	Private
Zhang et al. (75)	Cycle GAN	Private

utilized these two generators to generate images and masks from random noise.

Dirvanauskas et al. (63) generated human embryo cell images for three stages (one-cell, two-cell, and four-cell) by a conventional GAN model. All the synthesized images could be used to facilitate the development, training, and evaluation for embryo image processing tasks.

Zhang et al. (62) used GAN-based model to solve the data shortage problems in person re-identification task. Two view images (cross view images) are generated by a conditional GAN from existing original images and skeleton images. After that, these cross-view images are sent into a discriminator for person re-identification.

Frid-Adar et al. (65) used two variations of GAN models to generate synthetic liver lesions. The synthetic images contained the regions of interest (ROI) on abdomen CT images with a resolution of 64 × 64. Experiment results showed that the synthetic data augmentation from these two GAN models improved classification accuracy from 77.5 to 85.0% compared to the classic data augmentation.

Chen et al. (66) proposed a high-resolution MRI (HR MRI) image generation architecture. Instead of generating 2D HR MRIs, the authors generated 3D HR MRIs to learn 3D structures of MRI volumetric images. However, 3D networks bring more computing requirements. To solve this problem, the authors used 3D dense net-based architecture (67) in the generator. By combining 3D dense net and GAN, synthetic HR MRIs have more local image textures and details.

Mahapatra and Bozorgtabar (68) used local saliency maps and GAN for generating high-resolution retinal images. In addition to the GAN's lost function, they also added local saliency loss from the difference between HR images and low-resolution images in the saliency maps.

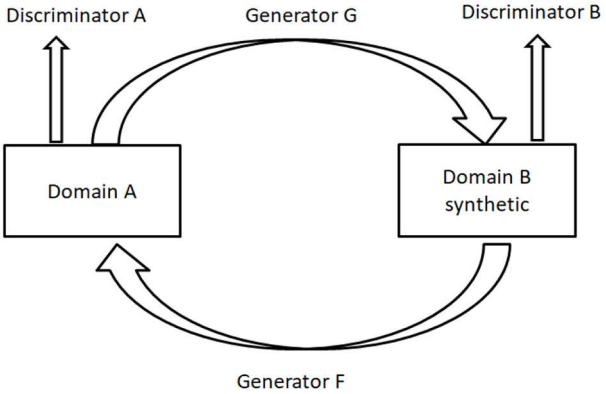

FIGURE 2 | The architecture of Cycle-GAN.

Yi and Babyn (69) proposed a deep neural network-based architecture for low dose CT denoizing. The generator-synthesized denoized CT image was sent to a sharpness detection network for comparison with a conventional CT image. This branch contributes a sharpness lost for the GAN objective function.

Shitrit and Raviv (70) applied GAN to accelerate the MRI image generation process. Instead of generating MRI images from existing MRI images, the authors used GAN to generate missing *k*-space samples. Their approach can be used for time-sensitive or resolution-sensitive MRI scan tasks.

Huo et al. (72) proposed a similar GAN architecture for splenomegaly segmentation. Instead of generating images from random noises, the authors used a U-Net (76) based architecture to get a segmented version and a Dice lost function as the discriminator. Isola et al. (77) used a Patch-GAN model as the discriminator for the patches from both the generated images and the ground truth images.

Cycle-GAN Used in Medical Image Analysis

Cycle-GAN (5) is utilized to learn the mapping from a domain image set X (or A in **Figure 2**) to another domain image set Y (or B in **Figure 2**) when the pairwise alignments between the two domains are unavailable. The forward generator is defined as G and the backward generator as F. The cycle consistency forces $F(G(x))\approx x$ and $G(F(y)) \approx$y. $F\circ G$ or $G\circ F$ works similarly to an auto-encoder (78) for learning the representations of the original images. A similar method was presented by Yi et al. (79).

It is essential that images from two different sources must have shared visual content, such as from wild horses to zebras. Normal image translation may be greatly affected when there is a significant difference between two domains. However, it is more suitable for multi-modality medical image synthesis, such as synthesizing images from Magnetic Resonance Imaging

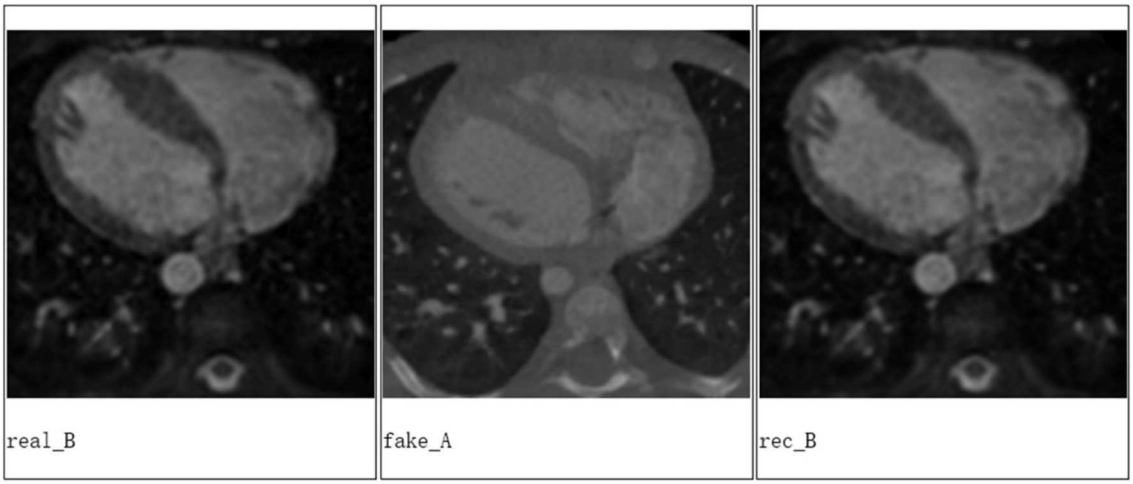

FIGURE 3 | The experiment results from Cycle-GAN, where real B is the real MRI image, fake A is the generated CT image based on the real MRI image, and rec B is the reconstructed MRI image based on the generated CT image of heart for a patient.

(MRI) to Computed Tomography (CT) (71) for diagnosing some specific diseases.

Wolterink et al. (71) applied Cycle-GAN to CT and MRI images of brain tumors in the radiotherapy treatment planning. Since a limited number of patients had paired CT and MRI images, unpaired images were generated from the paired images by first padding the images into a larger resolution and then cropping them into the same smaller resolution. Their results showed that Cycle-GAN trained with unpaired images outperforms a single GAN trained with paired images.

Hiasa et al. (4) improved the Cycle-GAN architecture by adding the gradient consistency loss with the goal of better edge alignment between the MRI images and CT images. Comparing the generated CT/MRI images with the actual ones, the gradient consistency had improved the synthesis accuracy and segmentation accuracy.

Huo et al. (80) combined Cycle-GAN and segmentation network in an end-to-end manner to take advantage of the complementary information between synthesis and segmentation. The final lost function of their network consisted of the Cycle-GAN lost functions and the segmentation loss function. Compared with the first synthesis and then segmentation method (81), the method from Huo et al. achieved better segmentation results on the spleen and other organs. Their experiments indicate that MRI images with multiple organ labels can be used to generate corresponding segmented CT images. Tanner et al. (73) used Cycle-GAN for MRI-CT deformable image registration of thoracic and abdominal organs. Jin et al. (82) applied Cycle-GAN for CT-MRI image synthesis. A discriminator was added for the real paired CT-MRI images and generated paired CT-MRI images. The combination of paired images and unpaired images achieved the lowest mean absolute error.

Yue et al. (74) proposed a task-driven generative model for X-ray image segmentation. A U-Net-based network (76) was trained supervised on Digitally Reconstructed Radiographs (DRRs) for organs segmentation. Thereafter, a Cycle-GAN was trained for DRRs and X-ray images synthesis. Specifically, the segmentation loss generated by the previously trained segmentation network was added to the cycle of real DRRs to fake X-ray to reconstructed DRRs. In that case, the segmentation results of X-ray images were greatly improved.

Zhang et al. (75) replaced discriminators with segmenters to address shape consistency problem. 3D fully convolution layers were used in the Cycle-GAN network and long-range U-Net network. Experiment results showed that the Cycle-GAN-based synthesis network and segmentation network were mutually beneficial in segmenting cardiovascular volumes.

We applied Cycle-GAN to achieve a good conversion between the CT and MRI images based on the data from MICCAI Workshop (**Figure 3**).

APPLICATION OF GAN IN MEDICAL INFORMATICS

With the development of health informatization, hospital information systems, Internet of Things (IoT)-based health Platform, wearable devices and other platforms have led to the explosive growth of medical data, such as electronic medical records (EMR) (83). The growth in the quantity and quality of medical data has also facilitated the use of scientific research and algorithms in medicine. However, due to data security, especially privacy security, although patient's data can be de-identified, the medical data after de-identification can still be re-identified by some combinations. Because there are barriers between the health information systems, it is very difficult to correlate medical data collected from different media, resulting in less medical data available for scientific research. The application of medical informatics often requires a large number of data

TABLE 3 | Literatures for the application of GAN in medical informatics.

References	Model	Public dataset
Choi et al. (84)	medGAN	PAMF, MIMIC III
Baowaly et al. (85)	medGAN, WGAN-GP, BGAN	MIMIC-III
Yoon et al. (86)	RadialGAN	MAGGIC
Che et al. (87)	ehrGAN	Private
Esteban et al. (88)	RGAN, RCGAN	Philips eICU database
Li et al. (89)	GAN	IQVIA longitudinal prescription (Rx) and medical claims (Dx) database
Guan et al. (90)	mtGAN	Private
Yang et al. (91)	GAN	UCI medical database, Cerebral stroke dataset
Tang et al. (92)	IRGAN	4705 hyperlipidemia questions from the internet
Hassouni et al. (93)	GAN	WISDM

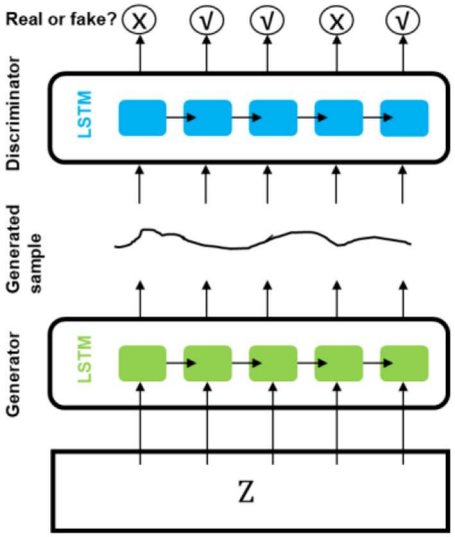

FIGURE 4 | The architecture of LSTM-based GAN in medical informatics.

to train parameters. The lack of medical data severely limits the application of deep learning algorithms, especially artificial intelligence in the field of medical informatics. Therefore, the development of medical informatics is behind fields such as medical images.

GAN has proven to play an important role in generating images and has shown good performance in generating continuous data. Since the gradient function is required to be differentiable, the traditional GAN cannot generate discrete data. In medical data, the diagnosis of a disease and the severity of the disease are discrete data. Due to the high cost, less labeled, and unbalanced classification medical data are available. Therefore, we explored the application of GAN in generating discrete data based on real medical data and in solving problems such as fewer labels and unbalanced classifications (**Table 3**).

Choi et al. (84) generated synthetic electronic health records (EHR) by using medical Generative Adversarial Networks (medGAN) based on the Sutter Palo Alto Medical Foundation (PAMF) and the Medical Information Mart for Intensive Care (MIMIC-III) datasets. The original GAN cannot be directly used to learn the discrete data of patients. medGAN can handle high-dimensional multi-label discrete variables (binary and count variables such as diagnoses, medications, and procedure codes) by leveraging the autoencoder to overcome the limitation from the original GAN. The autoencoder learned from real patient records and the same decoder in autoencoder was used to construct the discrete output after the generator. The authors obtained impressive results for discrete variables. Baowaly et al. (85) synthesized more realistic EHR than those generated by the medGAN using MIMIC-III, extended MIMIC-III, and Taiwan National Health Insurance Research Database (NHIRD). Two synthetic data generation models, Wasserstein GAN with gradient penalty (WGAN-GP) and boundary-seeking GAN (BGAN), were applied based on the medGAN framework. These two GAN models were named as medWGAN and medBGAN, respectively. The count (the frequency of a specific ICD or procedure of disease) and binary data (presence or absence of a specific ICD code) were created using medGAN, medWGAN, and medBGAN. Their results showed that the two improved GAN models outperformed the medGAN. medBGAN performed best in these three models.

Yoon et al. (86) used the auxiliary datasets, external datasets from related but different hospitals, as the noise for a GAN framework based on the fact that the patient distribution from one hospital will be better matched by the patient distribution from another hospital than by random noise such as Gaussian, enlarging the target dataset effectively. They used 14 studies of MAGGIC to create target datasets and compared the prediction performance between the proposed radialGAN, target-only GAN and benchmarks such as conditional-GAN and starGAN, etc. Integrating datasets from different hospitals by radialGAN can improve the performance of target-specific predictive models.

Che et al. (87) used two longitudinal real clinical datasets of heart failure and diabetes to investigate how well ehrGAN generated EHR as real samples. The structure of the basic prediction model was adopted in the discriminator. Based on the variational contrastive divergence, the generator was altered for semi-supervised learning setting. Data augmentation was performed by semi-supervised learning utilizing ehrGAN to boost risk prediction; thus, generalization capacity and prediction performance were improved.

Esteban et al. (88) presented a recurrent GAN (RGAN) and a recurrent conditional GAN (RCGAN) to generate sequences without/with some conditional inputs. Long Short-Term Memory (LSTM) was selected as the architecture for both discriminator and generator. They predicted whether or not a patient will become "critical" in the near future based on the four most frequently recorded variables measured by bedside monitors from Philips eICU database using a method named "Train on Synthetic, Test on Real" (TSTR). The models trained on the synthetic dataset from LSTM-based GAN (**Figure 4**) achieved performance at times comparable to that of the real data on the eICU dataset.

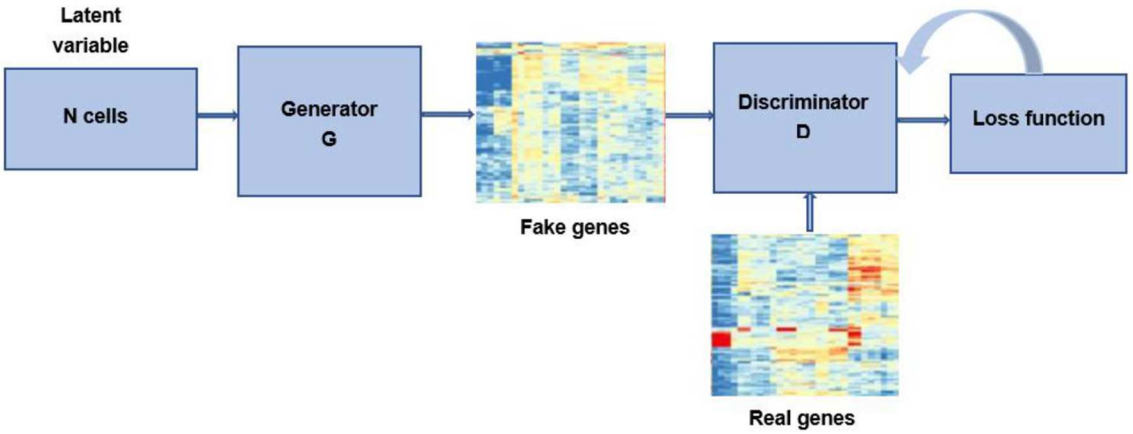

FIGURE 5 | Bioinformatics diagram of GAN.

Li et al. (89) used GAN to predict if a patient has a rare disease. The log probability of unlabeled data as real data was maximized and added in the objective function of the discriminator based on the IQVIA longitudinal prescription and medical claims database. Compared to the baseline techniques, the prediction accuracy of the semi-supervised learning framework for rare disease detection (precision-recall curves and area under the curve) was 5% higher.

Guan et al. (90) used medical text GAN (mtGAN) to reinforce the electronic medical record texts. mtGAN is a conditional GAN that takes designated disease features as input and generates corresponding EMR text to address privacy issues as well as inadequate and unbalanced and the insufficient and imbalance samples problem.

Yang et al. (91) presented a semi-supervised method in association with GAN to support medical decision making. In their study, GAN generated synthetic data by taking labeled set as input. Labels of the unlabeled set were predicted by two learners. By taking the expanded labeled set as input, GAN was used again to generate the labeled set. Both expanded labeled set and synthetic set were used as the training set to be classified based on the cerebral stroke set collected from IoT-based platform.

Tang et al. (92) proposed a GAN-based method to automatically retrieve patient questions. Supervised deep learning-based approaches were used to determine the similarity between patient questions. Their study showed that fine tuning with GAN can improve performance. Hassouni et al. (93) used GAN with LSTM to generate realistic simulation environments based on the WISDM dataset. Their results showed that the model trained on the data artificially generated by the GAN had similar performance trained on real data.

APPLICATION OF GAN IN BIOINFORMATICS

As a branch of the life science, bioinformatics is a new multidisciplinary field that understands and organizes information related to biomolecules through a combination of disciplines such as applied mathematics, biology, computer science, and statistics (94). It applies conventional statistics, modern computer science, machine learning, and other modeling algorithms to explore large volumes of biological data, including molecular sequences of DNA, RNA, proteins and metabolites, and other whole genome data. Bioinformatics research and applications include analysis of molecular sequence and genomics data; genome annotation; molecular folding, modeling, and design; building biological networks; analysis of the cellular organization and computational evolutionary biology (95, 96).

One of the most important and difficult issues for bioinformatics researchers is the accessibility and availability of large datasets. Though the increased throughput and technological advances have changed the landscape of metagenomics, the cost of sequencing is still relatively high. In addition, since the accessibility of data for research purpose involves many legal and ethical issues, bioinformatics data is highly sensitive (97). The lacking of available biological samples could result in imbalanced datasets, which can lead to over-fitting problems and poor classification performance. Recently, researches have used GAN to generate data samples to overcome these problems (**Figure 5**). Here, we present some of bioinformatics application cases (**Table 4**).

Ghahramani et al. (98) successfully simulated realistic scRNA-seq data using WGAN-GP, covering diverse scRNA-seq datasets of various cell types, including mouse epidermal, mouse neural and human hematopoietic single-cell RNA-seq data spanning from different laboratories and experimental protocols. The performance was evaluated at different checkpoints using *t*-distributed stochastic neighbor embedding (*t*-SNE) and the correlation between real cells and generated cells. As the training steps increased, the generator was capable of producing cells mapping to multiple clusters in the *t*-SNE plot, covering different cell types, cell states, cell origin and experimental batches present in the combined real dataset (98). By using the generative model, the researchers were able to obtain a universal representation

TABLE 4 | Literatures for the application of GAN in bioinformatics.

References	Model	Public dataset
Ghahramani et al. (98)	Wasserstein-GAN (WGAN) with gradient penalty loss function	GSE90848, GSE67602, GSE99989
Marouf et al. (99)	Conditional single-cell GAN	68,579 PBMCs (healthy donor A)
Xu et al. (6)	Generative adversarial networks for scRNA-seq imputation	GSE65525
Li et al. (100)	GAN	Not applicable
Anand and Huang (101)	GAN	Protein Data Bank
Killoran et al. (102)	GAN	Not applicable
Gupta and Zou (103)	Feedback GAN	Uniprot database
Wang et al. (104)	GAN	GEO, GTEx, 1000 G RNA-Seq expression data

of epidermal differentiation and the representation features can be used to predict the effect of cell state perturbations on gene expression at high temporal resolution. The task of the generator was to produce realistic output data from a random latent space vectors *z*. Corresponding latent space vectors of terminally differentiated and undifferentiated cells ($z_{differentiated}$and z_{basal}) were obtained according to the correlation between the real expression profiles and those generated, and the difference between these vectors was calculated as $\delta = z_{differentiated} - z_{basal}$. Then 1,000 time points were interpolated between the latent space differentiation vectors δ (98), so the dynamics of time-series gene expression over cell differentiation could be explored using the GAN model, which cannot be detected by experiments. By doing so, transiently expressed and subsequently down-regulated genes associated with the process of differentiation can be identified. They also performed a sensitivity analysis of the discriminator network to identify biological state-determining genes. By analyzing those networks, the authors obtained regulatory relationships for inferred genes.

Marouf et al. (99) built a single cell GAN model for scRNA-seq data generation using a Peripheral Blood Mononuclear Cell (PBMC) scRNA-seq dataset with 68,579 cells. A customed Library Size Normalization (LSN) function and Fully-Connected Neural Network with Batch Normalization were used in the scGAN's generator to improve training speed and stability. scGAN was able to model the dependency and correlation of intergenic, which are a hallmark of biological gene-regulatory networks. scGAN also captured gene count distributions and correlations well, and the training time was proportional to the complexity and size of scRNAseq datasets.

Marouf et al. (99) proposed a conditional scGAN model (cscGAN) for the realistic generation of single-cell RNA-seq data. The projection discriminator, along with the Conditional Batch Normalization and LSN function in the generator, is used to generate specific cell types of interest while learning multi-cell type complex data. Conditional generation of cell types could be used to augment the number of sparse specific cell populations that might represent only a small part of the total cells sequenced. However, it may help to solve the class imbalance problem. Similar to the above research, Wang et al. used CGAN for inferring target gene expression profiles by incorporating both adversarial and L1-norm loss terms. Comparative analysis showed that this model outperformed previous linear methods in gene expression inference (104).

Xu et al. proposed the GAN for scRNA-seq imputation (scIGANs), which uses generated realistic rather than observed cells to avoid the limitations, such as many sources of technical noises and dropouts, and the powerless for rare cells. ScIGANs converts the expression profiles of individual cells to images and feeds them to GAN. The trained generative network produces expression profiles representing the realistic cells of defined types. The generated cells, rather than the observed cells, are then used to impute the dropouts of the real cells (6).

Hi-C is commonly used to study three-dimensional genome organization. Hong et al. (105) developed a GAN, namely DeepHiC, to predict high-resolution Hi-C contact maps from low-coverage sequencing data. DeepHiC can reproduce high-resolution Hi-C data from as few as 1% down sampled reads. Application of DeepHiC to Hi-C data on mouse embryonic development can facilitate chromatin loop detection with higher accuracy.

Killoran et al. (102) developed a WGAN-based deep generative network for creating new DNA sequences by encoding the discrete sequences of characters (the nucleotides A, C, G, T) into a continuous representation using one-hot encodings. The authors proposed a joint approach that extends an activation maximization version by incorporating a trained generator model on a dataset of 4.6M 50-nucleotide-long sequences encompassing chromosome 1 of the human genome hg38. This approach is suitable for discrete sequences such as DNA. They found that the generative model can learn important structures from DNA sequences, and can be used to explore and design new DNA sequences with desired properties (102).

Guptaand Zou (103) proposed a novel feedback-loop architecture, called Feedback GAN (FBGAN) to optimize synthetic gene sequences for desired properties using an external function analyzer. The feedback-loop model consists of two components, including GAN and a differentiable neural network. GAN was used to generate novel raw gene sequences. The differentiable neural network named analyzer converted a gene sequence into a probability that the sequence encoded an antimicrobial peptide (AMP). The *n* top-ranked favorable generated sequences replaced the oldest *n* genes present in the discriminator training dataset. This model was able to generate synthetic genes coding for peptides of up to 50 amino acids in length, and the peptides can be optimized for the secondary alpha-helical structure of the resulting peptides (103).

Anand and Huang (101) applied GAN to generate protein structures by encoding protein structures in terms of pairwise distances between α-carbons on the protein backbone by using data from the Protein Data Bank, and used the Alternating Direction Method of Multipliers (ADMM) and Rosetta algorithm to transform 2D pairwise distances into 3D Cartesian coordinates. The authors compared their work with traditional HMMs-based methods. They found that their generator model could learn to construct meaningful secondary

structure elements such as alpha helices and beta sheets. The generated maps were highly variable and similar but not identical to the actual data, indicating that the GAN model was not just memorizing the training data. Finally, the authors verified that the generative model can reconstruct missing sections of corrupted protein structures (101).

Similarly, Yeh et al. (106) also proposed a 2D distance map representation of protein GAN model to predict particular missing regions in a protein structure using the idea of image inpainting. The author used this model to learn the distribution of a particular loop region with the context of the loop region from the candidate patch pool and successfully predicted the loop region (100). Compared to the traditional time-consuming and expensive experimental methods such as X-ray crystallography or Nuclear Magnetic Resonance (NMR), this GAN model is more convenient and time-saving.

Accurate identification of prognostic biomarkers is an important but challenging goal in bioinformatics. Kim et al. (107) applied GAN model to specify candidate prognostic gene module by graph learning algorithms and evaluated genes scores via a PageRank algorithm using multiple-omics data, including copy number, gene expression, DNA methylation, and somatic mutation data from five cancer types. Firstly, they reconstructed functional interaction networks (FIs network) that included known pathways in human biology. Then the reconstructed FIs network was learned by GAN to select features via PageRank with GAN weights, and finally, the prognosis was predicted. They successfully identified a number of genes involved in cancer development and analyzed their roles in biological pathways. Their model showed better predictive accuracy than existing methods.

There are some other applications in bioinformatics using GAN to enhance gene expression classification. Huynh et al. realized new data generation from original training datasets through the combination of GAN with nonlinear Support Vector Machines (SVMs). The results of GAN-SVM model displayed a better performance than the most advanced classification methods, including *k*-nearest neighbors (KNN), SVMs, decision tree (DTs) of C4.5, and random-forest (RFs) (108). Bhat et al. proposed a deep generative machine learning architecture (DeepCancer) to test the ability of GAN in classifying breast cancer and prostate cancer samples via the features learned by the discriminator. The results showed that the generative model achieved a high accuracy score (109).

CONCLUSION

In this article, we briefly introduced the origin, working principle, the development history of GAN, and numerous applications to the areas of digital image processing, medical imaging analysis, medical informatics, and bioinformatics.

In digital image processing, GAN can do image generation, high-resolution images generation from low-resolution images and image inpainting, which perform well and are widely used. Considering too many applications in digital image processing, due to space limitations, we only selected the articles with the most citations as the lead application to introduce, which leads to some applications in the directions of image processing weren't introduced comprehensively such as Style migration, image coloring, etc.

From the application of GAN in medical images, we can see that GAN-based models provide a good solution for data shortage in medical image analysis. It can be regarded as one of the important additions to the manual labeling from radiologists. The models based on a single GAN are more used as data augmentation methods to increase the variety and quantity of images in the same modality. On this basis, Cycle-GAN-based models make translations between multiple modalities possible. Involving in segmentation networks within the Cycle-GAN models, cross-modality segmentations can be learned in an end-to-end manner. This will effectively promote the application and development of deep learning algorithms in medical image analysis. However, GAN or Cycle-GAN still have limitations. For example, the CT images from the head cannot be generated from the CT images of the abdomen or MRI images from the legs. Researchers need to carefully design their data flow and lost functions to avoid problems such as non-convergence or model collapse. Additionally, some medical image analysis tasks require detailed 3D information of the organ. Involving 3D feature learning or 3D segmentation in the GAN-based model would be a challenge.

Based on EMR or EHR data, although GAN can generate realistic synthesized discrete data, continuous data, and even time series data to solve the issues of fewer labels and unbalanced classifications in medical informatics, there are still some limitations. How to evaluate these generated data and how to apply these generated data to solve medical problems has been controversial, which requires real data to validate.

So far, the application of GAN models in bioinformatics is still in a relatively early stage of development. Most studies applied GAN to generate and/or augment datasets. The above results have demonstrated the similarity of the data generated using GAN models to the original data. Most machine learning algorithms work well when the number of cases in each class is roughly equal. So, using the generated data, we can not only perform a lot of downstream analyses, such as detecting marker genes, dimensionality reduction and clustering, and reconstructing a particular secondary structure, but also decrease the number of human and animal experiments with a concomitant reduction in experimental costs, addressing important ethical and financial issues. In addition, GAN framework can work with any type of neural networks, so its application in bioinformatics will be more extensive.

In conclusion, more and more applications of GAN in biomedical research are being proposed. Some of GANs such as WGAN, Cycle-GAN, CGAN, medGAN are receiving more and more attention because of their importance in biomedical research. Although GAN has its advantages in simulating various problems, there are also some limitations. For example, when the sample size is small, the accuracy of the model will be relatively low. As GAN uses deep

neural networks as generators, poor interpretability is also a common problem.

AUTHOR CONTRIBUTIONS

LL, LY, ZZ, and ZF reviewed the literature and wrote the initial manuscript. LL, NZ, and XZ proposed an article outline. WZ, XZ, NZ, and YC revised the drafts. LL created and maintained an EndNote database.

ACKNOWLEDGMENTS

We thank Sichuan Science and Technology Program (No. 2019YFS0147), 1·3·5 project for disciplines of excellence, West China Hospital, Sichuan University (No. ZYJC18010), 1·3·5 project for disciplines of excellence-Clinical Research Incubation Project, West China Hospital, Sichuan University (No. 2019HXFH022), and the medical big data and statistical application technology workshop.

REFERENCES

1. Goodfellow IJ, Pouget-Abadie J, Mirza M, Bing X, Warde-Farley D, Ozair S, et al. Generative Adversarial Nets. In: *International Conference on Neural Information Processing Systems* Montréal, QC (2014).
2. Lin Y, Dai X, Li L, Wang X, Wang F. The new Frontier of AI research: generative adversarial networks. *Acta Autom Sin.* (2018) 44:775–92. doi: 10.16383/j.aas.2018.y000002
3. Zhang Z, Zhou X, Zhao S, Zhang X. Semantic prior guided face inpainting. In: Zhang Z, Zhou X, Zhao S, Zhang X, editors. *MMAsia '19*. New York, NY (2019).
4. Hiasa Y, Otake Y, Takao M, Matsuoka T, Sato Y. Cross-modality image synthesis from unpaired data using CycleGAN: effects of gradient consistency loss and training data size. In: Goksel O, Oguz I, Gooya A, Burgos N, editors. *Simulation and Synthesis in Medical Imaging - Third International Workshop, SASHIMI 2018, Held in Conjunction with MICCAI 2018, Proceedings*. Granada: Springer verlag (2018).
5. Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *IEEE International Conference on Computer Vision*. Venice (2017).
6. Xu Y, Zhang Z, You L, Liu J, Fan Z, Zhou X. Single-cell RNA-seq imputation using generative adversarial networks. *bioRxiv*. (2020). doi: 10.1101/2020.01.20.913384
7. Welander P, Karlsson S, Eklund A. Generative adversarial networks for image-to-image translation on multi-contrast MR images - a comparison of CycleGAN and unit. *arXiv* [Preprint]. (2018) arXiv: 1806.7777.
8. Li M, Tang H, Chan MD, Zhou X, Qian X, Li M. DC-AL GAN: Pseudoprogression and true tumor progression of glioblastoma multiform image classification based on DCGAN and AlexNet. *Med Phys.* (2019) 47, 1139–50. doi: 10.1002/mp.14003
9. Li H. *Statistical Learning Method*. Beijing: Tsinghua University Press (2012).
10. Feng X, Liu M, Liu J, Qin B, Sun Y, Liu T. Topic-to-essay generation with neural networks. In: Feng X, Liu M, Liu J, Qin B, Sun Y, Liu T, editors. *IJCAI*. Stockholm (2018). p. 4078–84. doi: 10.24963/ijcai.2018/567
11. Xian W, Xu K, Hall P. A survey of image synthesis and editing with generative adversarial networks. *Tsinghua Sci Technol.* (2017) 22:660–74. doi: 10.23919/TST.2017.8195348
12. Ledig C, Theis L, Huszar F, Caballero J, Aitken A, Tejani A, et al. Photo-realistic single image super-resolution using a generative adversarial network. *arXiv:1609.04802*. (2016). doi: 10.1109/CVPR.2017.19
13. Benaim S, Wolf L. One-sided unsupervised domain mapping. *arXiv:1706.00826*. (2017).
14. Yang Z, Yuan Y, Wu Y, Salakhutdinov R, Cohen WW. Review networks for caption generation. *arXiv:1605.07912* (2016).
15. Alom MZ, Taha TM, Yakopcic C, Westberg S, Sidike P, Nasrin MS, et al. A state-of-the-art survey on deep learning theory and architectures. *Electronics.* (2019) 8:292. doi: 10.3390/electronics8030292
16. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc.* (1977) 39:1–38. doi: 10.1111/j.2517-6161.1977.tb01600.x
17. Aliferis CF, Tsamardinos I, Statnikov A. HITON: a novel Markov Blanket algorithm for optimal variable selection. *AMIA Annu Symp Proc.* (2003) 2003:21–5.
18. Zeng N, Wang Z, Zhang H, Liu W, Alsaadi FE. Deep belief networks for quantitative analysis of a gold immunochromatographic strip. *Cogn Comput.* (2016) 8:684–92. doi: 10.1007/s12559-016-9404-x
19. Rasmussen CE. The infinite gaussian mixture model. In: *International Conference on Neural Information Processing Systems*. Breckenridge, CO (1999).
20. Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc IEEE.* (1989) 77:257-86. doi: 10.1109/5.18626
21. Goodfellow I. NIPS 2016 tutorial: generative adversarial networks. *arXiv:1701.00160* (2017).
22. Creswell A, White T, Dumoulin V, Kai A, Sengupta B, Bharath AA. Generative adversarial networks: an overview. *IEEE Signal Proc Mag.* (2018) 35:53–65. doi: 10.1109/MSP.2017.2765202
23. Spall JC. A stochastic approximation technique for generating maximum likelihood parameter estimates. In: *American Control Conference*. St. Louis, MO (2009).
24. Guo Y, Yu L, Oerlemans A, Lao S, Song W, Lew MS. Deep learning for visual understanding: a review. *Neurocomputing.* (2016) 187:27–48. doi: 10.1016/j.neucom.2015.09.116
25. Deng LY. The cross-entropy method: a unified approach to combinatorial optimization, Monte-Carlo simulation, and machine learning. *Technometrics.* (2004) 48:147–8. doi: 10.1198/tech.2006.s353
26. Arjovsky M, Chintala S, Bottou L. Wasserstein GAN. *arXiv:1701.07875* (2017).
27. Wang K, Chao G, Duan Y, Lin Y, Zheng X, Wang FY. Generative adversarial networks: introduction and outlook. *IEEE/CAA J Autom Sin.* (2017) 4:588–98. doi: 10.1109/JAS.2017.7510583
28. Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A, Xi C. Improved techniques for training GANs. *arXiv:1606.03498* (2016).
29. Hong Y, Hwang U, Yoo J, Yoon S. How generative adversarial networks and their variants work. *ACM Comput Surv.* (2019) 52:1–43. doi: 10.1145/3301282
30. Mirza M, Osindero S. Conditional generative adversarial nets. *arXiv [Preprint]*. (2014):2672–80. *arXiv*:1411.1784.
31. Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. *Comput Sci.* (2015).
32. Nowozin S, Cseke B, Tomioka R. f-GAN: training generative neural samplers using variational divergence minimization. *arXiv:1606.00709* (2016).
33. Fenchel W. On conjugate convex functions. In: Giorgi G, Kjeldsen T, editors, *Traces and Emergence of Nonlinear Programming*. Basel: Birkhäuser (2014).
34. Zhao J, Mathieu M, LeCun Y. Energy-based generative adversarial network. *arXiv:1609.03126* (2016).
35. Mao X, Li Q, Xie H, Lau RYK, Zhen W, Smolley SP. Least squares generative adversarial networks. *arXiv:1611.04076*. (2016). doi: 10.1109/ICCV.2017.304
36. Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville A. Improved training of wasserstein GANs. *arXiv:1704.00028* (2017).
37. Karras T, Aila T, Laine S, Lehtinen J. Progressive growing of GANs for improved quality, stability, and variation. *arXiv:1710.10196* (2017).
38. Pathak D, Krahenbuhl P, Donahue J, Darrell T, Efros AA. Context encoders: feature learning by inpainting. In: *IEEE Conference on Computer Vision and Pattern Recognition.* Las Vegas, NV: IEEE (2016).
39. Jin Y, Zhang J, Li M, Tian Y, Fang Z. Towards the automatic anime characters creation with generative adversarial networks. *arXiv:1708.05509* (2017).

40. Wang S, Sun J, Mehmood I, Pan C, Chen Y, Zhang YD. Cerebral micro-bleeding identification based on a nine-layer convolutional neural network with stochastic pooling. *Concurrency Comput Pract Exp.* (2020) 32:e5130. doi: 10.1002/cpe.5130
41. Jiang X, Zhang Y. Chinese sign language fingerspelling via six-layer convolutional neural network with leaky rectified linear units for therapy and rehabilitation. *J Med Imag Health Inform.* (2019) 9:2031–90. doi: 10.1166/jmihi.2019.2804
42. Li W, Jiang X, Sun W, Wang S, Liu C, Zhang X, et al. Gingivitis identification via multichannel gray-level co-occurrence matrix and particle swarm optimization neural network. *Int J Imag Syst Technol.* (2019) 1–11. doi: 10.1002/ima.22385
43. Wang S, Zhang Y, Yang M, Liu B, Ramirez J, Górriz JM. Unilateral sensorineural hearing loss identification based on double-density dual-tree complex wavelet transform and multinomial logistic regression. *Integr Comput Aided Eng.* (2019) 26:411–26. doi: 10.3233/ICA-190605
44. Zeng N, Wang Z, Zhang H, Kim K-E, Li Y, Liu X. An improved particle filter with a novel hybrid proposal distribution for quantitative analysis of gold immunochromatographic strips. *IEEE Trans Nanotechnol.* (2019) 18:819–29. doi: 10.1109/TNANO.2019.2932271
45. Zeng N, Wang Z, Zineddin B, Li Y, Du M, Xiao L, et al. Image-based quantitative analysis of gold immunochromatographic strip via cellular neural network approach. *IEEE Trans Med Imaging.* (2014) 33:1129–36. doi: 10.1109/TMI.2014.2305394
46. Zeng N, Zhang H, Li Y, Liang J, Dobaie AM. Denoising and deblurring gold immunochromatographic strip images via gradient projection algorithms. *Neurocomputing.* (2017) 247:165–72. doi: 10.1016/j.neucom.2017.03.056
47. Zeng N, Li H, Li Y, Luo X. Quantitative analysis of immunochromatographic strip based on convolutional neural network. *IEEE Access.* (2019) 7:16257–63. doi: 10.1109/ACCESS.2019.2893927
48. Zhang Y, Govindaraj V, Tang C, Zhu W, Sun J. High performance multiple sclerosis classification by data augmentation and AlexNet transfer learning model. *J Med Imag Health Inform.* (2019) 9:1–10. doi: 10.1166/jmihi.2019.2692
49. Zeng N, Zhang H, Song B, Liu W, Li Y, Dobaie AM. Facial expression recognition via learning deep sparse autoencoders. *Neurocomputing.* (2017) 273:643–49. doi: 10.1016/j.neucom.2017.08.043
50. Zhang G, You L, Lan L, Zeng N, Chen W, Poehling GG, et al. Risk prediction model for knee arthroplasty. *IEEE Access.* (2019) 7:34645–54. doi: 10.1109/ACCESS.2019.2900619
51. Szegedy C, Ioffe S, Vanhoucke V. Inception-v4, inception-ResNet and the impact of residual connections on learning. *arXiv:1602.07261* (2016).
52. Dai J, Yi L, He K, Jian S. R-FCN: Object Detection via region-based fully convolutional networks. *arXiv:1605.06409* (2016).
53. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, MA: IEEE (2015).
54. Shuiwang J, Ming Y, Kai Y. 3D convolutional neural networks for human action recognition. *IEEE Trans Pattern Anal Mach Intell.* (2013) 35:221–31. doi: 10.1109/TPAMI.2012.59
55. Kayalibay B, Jensen G, Smagt PVD. CNN-based segmentation of medical imaging data. *arXiv:1701.03056* (2017).
56. Li Q, Cai W, Wang X, Zhou Y, Feng DD, Chen M. Medical image classification with convolutional neural network. In: *International Conference on Control Automation Robotics and Vision.* IEEE: Singapore (2014).
57. Yitan Z, Peng Q, Yuan J. TCGA-assembler: open-source software for retrieving and processing TCGA data. *Nat Methods.* (2014) 11:599–600. doi: 10.1038/nmeth.2956
58. Deng J, Dong W, Socher R, Li LJ, Li K, Li FF. ImageNet: a large-scale hierarchical image database. In: *IEEE Conference on Computer Vision and Pattern Recognition*. Miami, FL (2009). doi: 10.1109/CVPR.2009.5206848
59. Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, et al. Microsoft COCO: common objects in context. *arXiv:1405.0312.* (2014). doi: 10.1007/978-3-319-10602-1_48
60. Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imaging.* (2016) 35:1285–98. doi: 10.1109/TMI.2016.2528162
61. Wang S-H, Xie S, Chen X, Guttery DS, Tang C, Sun J, et al. Alcoholism identification based on an AlexNet transfer learning model. *Front Psychiatry.* (2019) 10:205. doi: 10.3389/fpsyt.2019.00205
62. Zhang C, Zhu L, Zhang S, Yu W. PAC-GAN: an effective pose augmentation scheme for unsupervised cross-view person re-identification. *Neurocomputing.* (2019) 387:22–39. doi: 10.1016/j.neucom.2019.12.094
63. Dirvanauskas D, Maskeliunas R, Raudonis V, Damasevicius R, Scherer R. HEMIGEN: human embryo image generator based on Generative Adversarial Networks. *Sensors (Basel).* (2019) 19:3578. doi: 10.3390/s19163578
64. Pandey S, Singh PR, Tian J. An image augmentation approach using two-stage generative adversarial network for nuclei image segmentation. *Biomed Signal Process.* (2020) 57:101782. doi: 10.1016/j.bspc.2019.101782
65. Frid-Adar M, Diamant I, Klang E, Amitai M, Goldberger J, Greenspan H. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing.* (2018) 321:321–31. doi: 10.1016/j.neucom.2018.09.013
66. Chen Y, Feng S, Christodoulou AG, Zhou Z, Li D. Efficient and accurate MRI super-resolution using a generative adversarial network and 3D multi-level densely connected network. *arXiv:1803.01417.* (2018). doi: 10.1007/978-3-030-00928-1_11
67. Chen Y, Xie Y, Zhou Z, Feng S, Christodoulou AG, Li D. Brain MRI super resolution using 3D deep densely connected neural networks. In: *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018).* IEEE: Washington, DC (2018). pp. 739–42.
68. Mahapatra D. Retinal vasculature segmentation using local saliency maps and Generative adversarial networks for image super resolution. *arXiv:1710.04783* (2017).
69. Yi X, Babyn P. Sharpness-aware low-dose CT denoising using conditional Generative Adversarial Network. *J Digit Imaging.* (2018) 31:1–15. doi: 10.1007/s10278-018-0056-0
70. Shitrit O, Raviv TR. Accelerated magnetic resonance imaging by adversarial neural network. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017*. Québec, QC (2017). doi: 10.1007/978-3-319-67558-9_4
71. Wolterink JM, Dinkla AM, Savenije MHF, Seevinck PR, Berg CATV, Išgum I. Deep MR to CT synthesis using unpaired data. *arXiv:1708.01155*. (2017). doi: 10.1007/978-3-319-68127-6_2
72. Huo Y, Xu Z, Bao S, Bermudez C, Moon H, Parvathaneni P, et al. Splenomegaly segmentation on multi-modal MRI using Deep Convolutional Networks. *IEEE Trans Med Imaging.* (2019) 38:1185–96. doi: 10.1109/TMI.2018.2881110
73. Tanner C, Ozdemir F, Profanter R, Vishnevsky V, Konukoglu E, Goksel O. Generative Adversarial Networks for MR-CT deformable image registration. *arXiv:1807.07349* (2018).
74. Yue Z, Miao S, Mansi T, Rui L. Task driven generative modeling for unsupervised domain adaptation: application to x-ray image segmentation. *arXiv:1806.07201* (2018).
75. Zhang Z, Lin Y, Zheng Y. Translating and segmenting multimodal medical volumes with cycle- and shape-consistency Generative Adversarial Network. *arXiv:1802.09655*. (2018). doi: 10.1109/CVPR.2018.00963
76. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-assisted Intervention.* Cham (2015).
77. Isola P, Zhu JY, Zhou T, Efros AA. Image-to-Image translation with Conditional Adversarial Networks. *arXiv:1611.07004*. (2016). doi: 10.1109/CVPR.2017.632
78. Lu X, Tsao Y, Matsuda S, Hori C. Speech enhancement based on deep denoising Auto-Encoder. In: *Proceedings of Interspeech*. Lyon (2013). pp. 436–40.
79. Yi Z, Hao Z, Gong PTM, Yi Z, Hao Z, Gong PTM, et al. DualGAN: unsupervised dual learning for image-to-image translation. *arXiv:1704.02510*. (2017). doi: 10.1109/ICCV.2017.310

80. Huo Y, Xu Z, Bao S, Assad A, Abramson RG, Landman BA. Adversarial synthesis learning enables segmentation without target modality ground truth. *arXiv:1712.07695*. (2017). doi: 10.1109/ISBI.2018.8363790
81. Chartsias A, Joyce T, Dharmakumar R, Tsaftaris SA. Adversarial image synthesis for unpaired multi-modal cardiac data. In: Tsaftaris S, Gooya A, Frangi A, Prince J, editors. *Simulation and Synthesis in Medical Imaging. SASHIMI 2017. Lecture Notes in Computer Science*. Cham: Springer (2017).
82. Jin CB, Jung W, Joo S, Park E, Cui X. Deep CT to MR synthesis using paired and unpaired data. *arXiv:1805.10790*. (2018). doi: 10.3390/s19102361
83. Li Q, Lan L, Zeng N, You L, Yin J, Zhou X, et al. A framework for big data governance to advance RHINs: a case study of China. *IEEE Access.* (2019) 7:50330–8. doi: 10.1109/ACCESS.2019.2910838
84. Choi E, Biswal S, Malin B, Duke J, Stewart WF, Sun J. Generating multi-label discrete patient records using Generative Adversarial Networks. *arXiv:1703.06490*. (2017).
85. Baowaly MK, Lin C, Liu C, Chen K. Synthesizing electronic health records using improved generative adversarial networks. *J Am Med Inform Assoc.* (2019) 26:228–41. doi: 10.1093/jamia/ocy142
86. Yoon J, Jordon J, van der Schaar M. RadialGAN: leveraging multiple datasets to improve target-specific predictive models using Generative Adversarial Networks. *arXiv:1802.06403* (2018).
87. Che Z, Yu C, Zhai S, Sun Z, Yan L, Che Z, et al. Boosting deep learning risk prediction with Generative Adversarial Networks for electronic health records. *arXiv:1709.01648*. (2017). doi: 10.1109/ICDM.2017.93
88. Esteban C, Hyland SL, Rätsch G. Real-valued (Medical) time series generation with recurrent conditional GANs. *arXiv:1706.02633* (2017).
89. Li W, Wang Y, Cai Y, Arnold C, Zhao E, Yuan Y. Semi-supervised rare disease detection using generative adversarial network. *arXiv:1812.00547* (2018).
90. Guan J, Li R, Yu S, Zhang X. Generation of synthetic electronic medical record text. *arXiv:1812.02793v1*. (2018). pp. 374–80. doi: 10.1109/BIBM.2018.8621223
91. Yang Y, Nan F, Yang P, Meng Q, Xie Y, Zhang D, et al. GAN-based semi-supervised learning approach for clinical decision support in health-IoT platform. *IEEE Access.* (2019) 7:8048–57. doi: 10.1109/ACCESS.2018.2888816
92. Tang G, Ni Y, Wang K, Yong Q. Fine-tuning neural patient question retrieval model with Generative Adversarial Networks. *Stud Health Technol Inform.* (2018) 247:720–4.
93. Hassouni AE, Hoogendoorn M, Muhonen V. Using generative adversarial networks to develop a realistic human behavior simulator. In: Miller T, Oren N, Sakurai Y, Noda I, Savarimuthu B, Cao Son T, editors. *PRIMA 2018: Principles and Practice of Multi-Agent Systems. PRIMA 2018. Lecture Notes in Computer Science.* Cham: Springer (2018).
94. Shahnaz M, Chowdhary M, Asha Rani D, Parkash J, Dev D. Bioinformatics: an overview for cancer research. *J Drug Deliv. Ther.* (2016) 6:69–72. doi: 10.22270/jddt.v6i4.1290
95. Mount DW. *Bioinformatics: Sequence and Genome Analysis.* 2nd edn. New York, NY: Cold Spring Harbor Laboratory Press (2004).
96. Abdurakhmonov IY. Bioinformatics: basics, development, and future. *IntechOpen.* (2016) 3–26. doi: 10.5772/63817
97. Sharmilan S. Generate bioinformatics data using Generative Adversarial Network: a review. In: *2nd International Conference on Information Technology Research*. Colombo (2017).
98. Ghahramani A, Watt FM, Luscombe NM. Generative adversarial networks simulate gene expression and predict perturbations in single cells. *bioRxiv*. (2018) 262501. doi: 10.1101/262501
99. Marouf M, Machart P, Bansal V, Kilian C, Magruder DS, Krebs CF, et al. Realistic in silico generation and augmentation of single-cell RNA-seq data using generative adversarial networks. *Nat Commun.* (2020) 11:1–12. doi: 10.1038/s41467-019-14018-z
100. Li Z, Nguyen SP, Dong X, Yi S. Protein Loop Modeling Using Deep Generative Adversarial Network. In: *IEEE International Conference on Tools with Artificial Intelligence*. Boston, MA (2017).
101. Anand N, Huang P. Generative modeling for protein structures. In: *ICLR 2018 Workshop*. Vancouver, BC (2018).
102. Killoran N, Lee LJ, Delong A, Duvenaud D, Frey BJ. Generating and designing DNA with deep generative models. *arXiv:1712.06148* (2017).
103. Gupta A, Zou J. Feedback GAN (FBGAN) for DNA: a novel feedback-loop architecture for optimizing protein functions. *arXiv:1804.01694*. (2018).
104. Wang X, Ghasedi Dizaji K, Huang H. Conditional generative adversarial network for gene expression inference. *Bioinformatics*. (2018) 34:i603–11. doi: 10.1093/bioinformatics/bty563
105. Hong H, Jiang S, Li H, Du G, Sun Y, Tao H, et al. DeepHiC: A generative adversarial network for enhancing Hi-C data resolution. *Plos Comput Biol.* (2020) 16:e1007287. doi: 10.1371/journal.pcbi.1007287
106. Yeh RA, Chen C, Lim TY, Schwing AG, Hasegawajohnson M, Do MN. Semantic image inpainting with deep generative models. *arXiv*. (2016). doi: 10.1109/CVPR.2017.728
107. Kim M, Oh I, Ahn J. An improved method for prediction of cancer prognosis by network learning. *Genes.* (2018) 9:478. doi: 10.3390/genes9100478
108. Huynh P, Van Hoa Nguyen, Do T. Enhancing gene expression classification of support vector machines with generative adversarial networks. *KCI Accredited J.* (2019) 17:14–20.
109. Bhat RR, Viswanath V, Li X. DeepCancer: detecting cancer through gene expressions via deep generative learnin. *arXiv* (2016).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Public Health

ORIGINAL RESEARCH
published: 16 April 2021
doi: 10.3389/fpubh.2021.615597

Auxiliary Diagnostic Method for Patellofemoral Pain Syndrome Based on One-Dimensional Convolutional Neural Network

Wuxiang Shi[1,2], *Yurong Li*[2], *Dujian Xu*[3], *Chen Lin*[1,2], *Junlin Lan*[1,2], *Yuanbo Zhou*[1,2], *Qian Zhang*[1,2], *Baoping Xiong*[1,4]* *and Min Du*[1,5]*

[1] College of Physics and Information Engineering, Fuzhou University, Fuzhou, China, [2] Fujian Key Laboratory of Medical Instrumentation & Pharmaceutical Technology, Fuzhou University, Fuzhou, China, [3] Yida Equity Investment Fund Management Co., Ltd., Nanjing, China, [4] Department of Mathematics and Physics, Fujian University of Technology, Fuzhou, China, [5] Fujian Provincial Key Laboratory of Eco-Industrial Green Technology, Wuyi University, Wuyishan, China

Early accurate diagnosis of patellofemoral pain syndrome (PFPS) is important to prevent the further development of the disease. However, traditional diagnostic methods for PFPS mostly rely on the subjective experience of doctors and subjective feelings of the patient, which do not have an accurate-unified standard, and the clinical accuracy is not high. With the development of artificial intelligence technology, artificial neural networks are increasingly applied in medical treatment to assist doctors in diagnosis, but selecting a suitable neural network model must be considered. In this paper, an intelligent diagnostic method for PFPS was proposed on the basis of a one-dimensional convolutional neural network (1D CNN), which used surface electromyography (sEMG) signals and lower limb joint angles as inputs, and discussed the model from three aspects, namely, accuracy, interpretability, and practicability. This article utilized the running and walking data of 41 subjects at their selected speed, including 26 PFPS patients (16 females and 10 males) and 16 painless controls (8 females and 7 males). In the proposed method, the knee flexion angle, hip flexion angle, ankle dorsiflexion angle, and sEMG signals of the seven muscles around the knee of three different data sets (walking data set, running data set, and walking and running mixed data set) were used as input of the 1D CNN. Focal loss function was introduced to the network to solve the problem of imbalance between positive and negative samples in the data set and make the network focus on learning the difficult-to-predict samples. Meanwhile, the attention mechanism was added to the network to observe the dimension feature that the network pays more attention to, thereby increasing the interpretability of the model. Finally, the depth features extracted by 1D CNN were combined with the traditional gender features to improve the accuracy of the model. After verification, the 1D CNN had the best performance on the running data set (accuracy = 92.4%, sensitivity = 97%, specificity = 84%). Compared with other methods, this method could provide new ideas for the development of models that assisted doctors in diagnosing PFPS without using complex biomechanical modeling and with high objective accuracy.

Keywords: patellofemoral pain syndrome, one-dimensional convolutional neural network, focal loss, attention mechanism, joint angles, surface electromyography

OPEN ACCESS

Edited by:
Shuihua Wang,
University of Leicester,
United Kingdom

Reviewed by:
Siyuan Lu,
University of Leicester,
United Kingdom
Lin Wang,
Chinese Academy of Sciences (CAS), China

***Correspondence:**
Baoping Xiong
xiongbp@fjut.edu.cn
Min Du
dm_dj90@163.com

Specialty section:
This article was submitted to Digital Public Health, a section of the journal Frontiers in Public Health

Received: *09 October 2020*
Accepted: *04 March 2021*
Published: *16 April 2021*

Citation:
Shi W, Li Y, Xu D, Lin C, Lan J, Zhou Y, Zhang Q, Xiong B and Du M (2021) Auxiliary Diagnostic Method for Patellofemoral Pain Syndrome Based on One-Dimensional Convolutional Neural Network. Front. Public Health 9:615597. doi: 10.3389/fpubh.2021.615597

INTRODUCTION

Patellofemoral pain syndrome (PFPS) is a common knee joint disease in clinical practice, with a prevalence of 10–28% in the general population, about a quarter of the total population, which is often caused by degenerative changes of articular cartilage (1–3). This disease is common in athletes and women, causing severe pain during sports and daily activities, and it affects athletes' careers to a large extent (1, 4). PFPS will have a certain impact on the physical and mental health of patients, making the patients unable to lead an active life (5). Most of the daily activities, that is, up and down stairs, sitting, and squatting, will aggravate the pain of the patients (6). Moreover, PFPS may develop into patellofemoral osteoarthritis (7).

Timely detection and definite diagnosis are the keys to prevent the aggravation of PFPS, but they are not easy (2, 8). Despite the high incidence of PFPS, the pathophysiology of PFPS is unclear (9, 10). Considering that the onset of PFPS is caused by many factors, misjudgment easily occurs (11). At present, the cause of PFPS has two explanations. One is biomechanical joint dislocation, muscle weakness, and excessive joint load around the patella, and the other is pain caused by nerve structure on neurodynamic (6). According to the survey, no clear diagnostic criteria are available at present, but some acceptable reference standards are identified, such as patellar apprehension, patella palpation, patellar apprehension, Waldron test, compression test, and patellar tracking (2). However, these standards are mostly dependent on the subjective judgment of doctors, and the whole diagnosis results and medical effect are strongly related to the rich experience and knowledge of experts, which are not friendly to young doctors. Different standards will lead to different diagnosis results, and no accurate and unified standard is identified for judging PFPS; thus, the diagnostic accuracy is relatively poor (8, 12). Although some PFPS diagnoses in the form of questionnaires (such as the Kujala score) have high sensitivity and specificity, they rely on the subjective answers of the patient and include a certain degree of privacy of the patient, which is difficult for some patients to cooperate (13). At present, invasive or minimally invasive methods are primarily used to assist in the detection of knee injury and diseases. Among the methods, MRI, CT, and other non-invasive detection methods can be more effective in the detection of knee injury and diseases, but these large-scale instruments and equipment are expensive, which are not convenient for daily inspection. As a minimally invasive method, arthroscopy can provide detailed diagnosis information, but repeated incision of the knee joint will cause pain to patients, which is not conducive to the recovery of injury and diseases. Therefore, exploring a new high-precision and low-cost non-invasive PFPS detection method is necessary.

In recent years, increasing studies have focused on the relationship between PFPS and biomechanical parameters (2, 14, 15). Ferrari et al. used the mid-band parameters of surface electromyography (sEMG) to distinguish PFPS by independent *t*-test and other methods (2). Bernard et al. explored whether the coordination of body strength in patients with PFPS has changed (16). Besier et al. used electromyography and lower limb kinematics data to drive a musculoskeletal model and evaluate the muscle strength of PFPS patients and painless subjects during walking and running (17). Myer et al. used a multiple linear logistic regression model to predict the knee-abduction moment when athletes land and explore the relationship between high knee-abduction moment and increased risk of PFPS (18). However, most of the parameters required in these studies are obtained through artificial extraction or the biomechanical model, which is time-consuming. The biomechanical model is based on the musculoskeletal model to establish the relationship between the sEMG signal and joint movement. Nevertheless, the coordination mechanism of the human nerves, muscles, and skeletal system cannot be fully understood, which leads to the inability to accurately simulate the human neuromusculoskeletal system, which causes a fatal flaw in the calculation model, that is, an "individual error."

Previous studies have shown that when the principle of the system is not clear or unknown, the artificial neural network driven by data has good system characterization and individual adaptability (19). With the development of artificial intelligence technology, artificial neural network methods have been increasingly used in the field of biomechanics and disease diagnosis (20–22). For example, Keijsers et al. used plantar pressure measurements as input to an artificial neural network to classify forefoot pain (23). Otag et al. used an artificial neural network to obtain the ligamentum patellae angle and explained that the prevalence of PFPS in women is greater than that in men based on the difference in angle values between men and women. However, the accuracy in the classification of the left and right knees is mediocre, only 67% (24). Biomechanics will include a variety of non-linear problems, which can be well-solved by an artificial neural network. Thus, this study aims to construct a convolutional neural network (CNN) model to distinguish PFPS through several easy-to-measure biomechanical parameters. Traditional CNN mostly uses two-dimensional convolution, but these biomechanical parameters are generally time series, which have a certain periodicity; thus, this paper proposes to use one-dimensional convolution, causing the filters to only slide on the time axis. Retaining the correlation among various parameters can achieve the time variability of biomechanical parameters and improve the accuracy of network discrimination.

The main contribution of this study is to propose a high-precision, low-cost and easy-to-implement computer-aided diagnostic method, which provides a new idea for the development of a convenient PFPS diagnostic model. The focal loss function is introduced to optimize the network parameters, which improves the balance of the 1D CNN results. By adding attention mechanism into the network and visualizing the output features, we can increase the interpretability of the model to analyze the diversity of biomechanical features involved in PFPS. Moreover, some studies have shown that there are gender differences in PFPS. In this paper, the depth features extracted by one-dimensional CNN are combined with the traditional gender features, and these features are classified through the full connection layer to improve the accuracy of the model.

The rest of this paper is as follows. The second section introduces the data sets and preprocessing methods used in this experiment, and then introduces the neural network model

TABLE 1 | Mean ± SD age, height, and body mass of subjects.

	PFPS		Controls	
	Males (n = 10)	**Females (n = 16)**	**Males (n = 7)**	**Females (n = 8)**
Age (years)	30.5 ± 4.5	28.7 ± 4.6	27.2 ± 3.0	28.8 ± 4.7
Height (m)	1.78 ± 0.08	1.68 ± 0.06	1.80 ± 0.05	1.66 ± 0.05
Mass (kg)	73.5 ± 15.7	62.7 ± 10.0	73.4 ± 18.1	58.3 ± 4.6

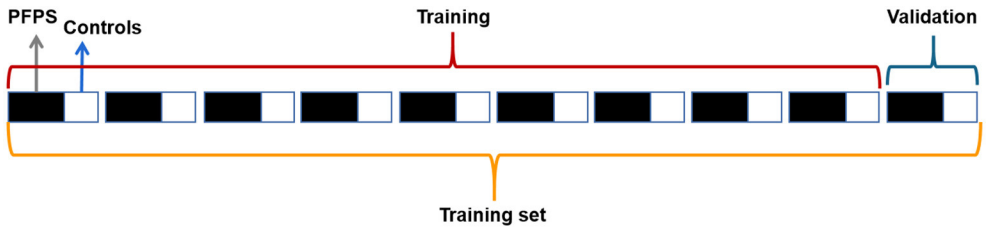

FIGURE 1 | Data partition in 10-fold cross-validation.

used in this experiment and the experimental environment in detail. In the third section, the experimental results are given and compared. The fourth section discusses the experimental results, and the fifth section summarizes and prospects the full text.

METHODS

Experimental Data

This study was a retrospective exploratory secondary analysis of a subset of an open data set. This public data set primarily recorded the lower limb kinematic data and sEMG signals of PFPS patients and painless control subjects during walking and running and muscle strength obtained from the musculoskeletal model (17). A total of 27 patients with patellofemoral pain (16 female, 11 male) and 16 painless control groups (eight female, eight male) were included in the study. These patients and painless controls were identified by professional doctors, and they were tested for walking, running, and squatting at a self-selected pace. In this paper, 10 kinds of biomechanical characteristics were selected in walking and running tests, which included three kinds of joint angle values [knee flexion (KF) angle, hip flexion (HF) angle, ankle dorsiflexion (ADF) angle], and seven kinds of sEMG signals [semimembranosus (SEB), rectus femoris (RF), biceps femoris short head (BF), vastus medialis (VM), vastus lateralis (VL), lateral gastrocnemius (LG), and medial gastrocnemius (MG)]. These parameters were selected because they were related to PFPS, which could be measured in real-time without using biomechanical modeling. The original sEMG data used a zero-lag fourth-order recursive Butterworth filter (30 Hz) for high-pass filtering and a Butterworth low-pass filter (6 Hz) for full-wave rectification and filtering. The detailed collection of the entire data set could be found in Reference (17). The experimental data used in this research were obtained from the public data set of this website (https://www.sciencedirect.com/science/article/pii/S0021929009000396?via%3Dihub).

Data Pre-processing

The data should be cleaned before placing into the neural network. Considering that certain data were missing in the walking and running data of subjects 4 and 43, we eliminated them and tested the data of the remaining 41 subjects, including 26 PFPS patients (16 female, 10 male) and 15 painless controls (eight female, seven male). Each subject had walking and running test data. We combined the data of each subject into a 100 * 10 matrix to adapt to the input form of a convolutional neural network (100 time-series recorded values, 10 characteristics). The relevant information on subjects is shown in **Table 1.**

The original data had already filtered out the noise, and no filter was needed, but we needed to standardize the parameters of each subject. The range of the joint angle value and EMG signal value was quite different, which was not conducive to the convergence of the neural network; thus, we standardized the range to make it consistent:

$$X_{\mathrm{i}}=\frac{X_{\mathrm{i}}-\bar{X}}{X_{\mathrm{std}}}, \tag{1}$$

where $\bar{X}$ is the mean of each feature of the original data X, and X_{std} is the variance of each feature of the original data X.

The preprocessed data were equivalent to a two-dimensional matrix. We flipped the data in the training set horizontally, but we cannot flip such date vertically because the column represented the time axis, which had strong correlation. Therefore, the number of training sets can be doubled, and the performance of the neural network model can be improved.

Experimental Protocol

We randomly selected 70% of the subjects as the training set and 30% as the test set, and the proportion of PFPS patients and painless controls in the training set was the same as that in the test set. The training set and test set were processed similarly, and then the training set was placed into the neural network

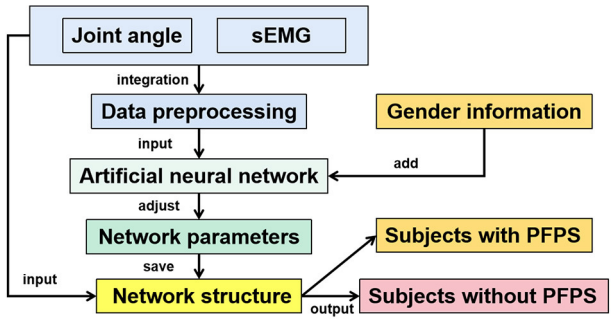

FIGURE 2 | Overall flow chart of the method.

for training. Considering that our data set was small and the proportion of PFPS patients was large, we adopted hierarchical 10-fold cross-validation to adjust the network parameters, avoid specificity, and maximize the utilization of data. The training set was equally divided into 10 equal parts, and the proportional relationship between PFPS patients and painless controls in each set was the same. Nine of them were used to train the network, and one was used for verification, which was circularly repeated 10 times to ensure that each copy was used, which is shown in **Figure 1**.

In this paper, several artificial neural network models commonly used in classification tasks were selected for testing, including extreme learning machine (ELM), back propagation neural network (BP), one-dimensional convolution neural network (1D CNN), two-dimensional convolution neural network (2D CNN), long short-term memory (LSTM), VGGNet, and AlexNet. The BP neural network here refers to a fully connected neural network with a hidden layer. This article focused on the 1D CNN, and the other neural networks were primarily used for comparison. Except for VGGNet and AlexNet, all parameters of other artificial neural networks were obtained through 10-fold cross-validation to avoid particularity. The overall flow chart of the method is shown in **Figure 2**.

Network Structure

CNN has been proven to have great advantages in a variety of classification tasks, such as image recognition, natural language processing, and human action recognition (25–30). In recent years, a number of excellent CNN classification models have been created, such as AlexNet (31) and VGGNet (32). These two network models belong to the best of their kind, particularly in image classification. In addition, they are often found in medical image classification, which is a good computer-aided diagnostic method. These two models have many parameters. For small data sets, most researchers use transfer learning (33, 34). The data set in this paper is also relatively small, but it is not suitable for transfer learning, because the premise of transfer learning is that the data in the original task and the target task are similar, that is, there is a certain Association for learning. However, most of the training data used in these large-scale classification models such as AlexNet and VGGNet are based on image data, which is very different from the multidimensional time series data in this paper, so it is not applicable.

Most of the CNN convolution kernels are two-dimensional. However, according to the characteristics of biomechanical parameters belonging to time-series data, this article utilized the 1D CNN for learning. The network structure of 1D CNN in this paper is shown in **Figure 3**. We replaced the convolution kernel in the AlexNet model and VGGNet model with one-dimensional convolution kernel to make a better comparison, and other network structures remained unchanged.

Our inputs were the 100 * 10 matrixes. First, we added a soft attention mechanism to the input, which could reweight the input information adaptively before convolution. This process separated important input features. Then, in the first convolutional layer, we defined 16 filters (also known as feature detector) with the convolution kernel size of 3. The filters only slid on the time axis, and the sliding step size was 1. During training of the first layer, we obtained 16 different feature maps. The structure of the filters in the second convolutional layer was the same as that of the first layer, which was used to learn complex features. The max pooling layer would slide a window of height 2 on the feature map with a step size of 1 and replace it with the maximum value, which discarded half of the value. After the pooling operation, part of the information would be lost; thus, the number of filters in the next two convolutional layers was increased to 32. We added a dropout layer with a dropout ratio of 0.3 (30% of neurons were randomly ignored) after the last convolutional layer to avoid overfitting. Then, we expanded the feature map output of the convolution layer into a one-dimensional vector. Simultaneously, we placed the gender characteristics through binary encoding (01 for males and 10 on behalf of females) and fused such characteristics with the depth feature extracted from the convolution layer. Finally, the fused features were placed into a fully connected neural network with 50 neurons for learning, which were reduced to a vector of length 2 (representing the two types of output) through the softmax activation function. Meanwhile, the optimization algorithm selected Adam and set the learning rate to 0.00001 and the number of iterations to 4,000.

The network structure of the 2D CNN was similar to that of the 1D CNN; however, the convolution kernels of the 2D CNN were two-dimensional, which were set to 3 * 3. This network was designed to facilitate comparison with the 1D CNN. The network structure of ELM and BP only had a single hidden layer. The number of neurons in the hidden layer of ELM and BP was 174 and 37, respectively, which were obtained by ten-fold cross-validation (**Figure 4**).

In addition to ELM, other neural networks optimized the parameters by reducing loss. The ordinary cross-entropy loss function was used to optimize the network parameters in most artificial neural networks. Given the large proportion of PFPS tags in the data set, misjudging painless subjects as PFPS by the neural network was easy. Thus, we utilized the focal loss function, which could solve the problem of imbalance between positive and negative samples and reduce the impact of easy-to-predict samples

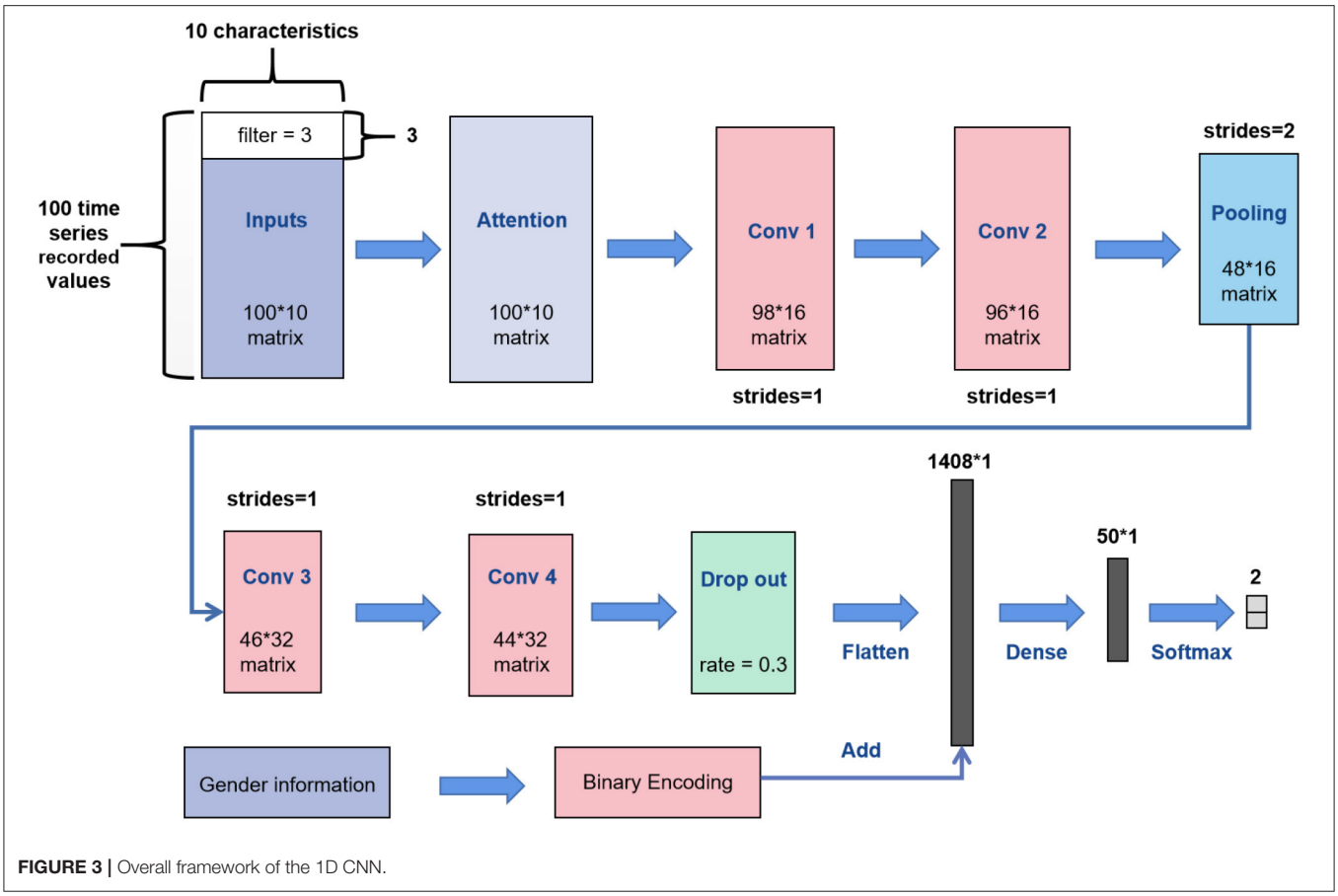

FIGURE 3 | Overall framework of the 1D CNN.

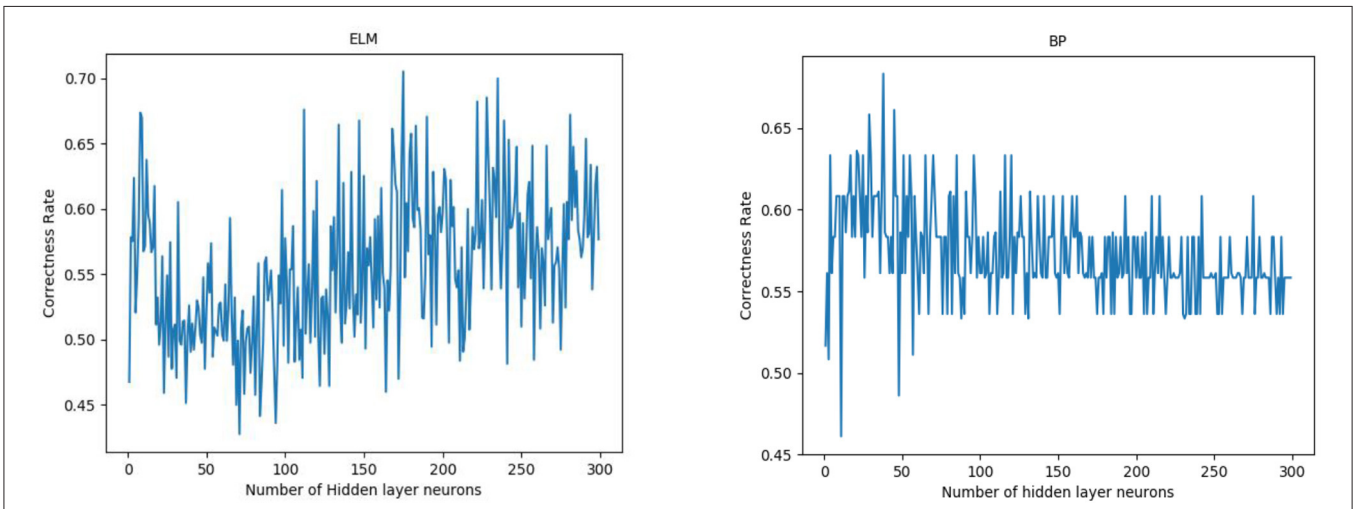

FIGURE 4 | From left to right are the 10-fold cross-validation results of ELM and BP on the running dataset.

(35, 36):

$$\mathbf{LOSS} = -\mathbf{a}(\mathbf{1}-\mathbf{y}')^{\mathbf{r}*}\mathbf{logy}', \qquad \mathbf{y}=\mathbf{1}, \tag{2}$$

$$\mathbf{LOSS} = -(\mathbf{1}-\mathbf{a})\mathbf{y}'^{\mathbf{r}*}\mathbf{log}(\mathbf{1}-\mathbf{y}'), \qquad \mathbf{y}=\mathbf{0}, \tag{3}$$

where $y = 1$ is the label of PFPS, and $y = 0$ is the label of painless control. y' is the corresponding predicted label. α is the balance adjustment factor, and r is used to control the rate of adjustment. When the sample is easy to predict, that is, y' is larger, its weight $1 - y'$ will be smaller. Meanwhile,

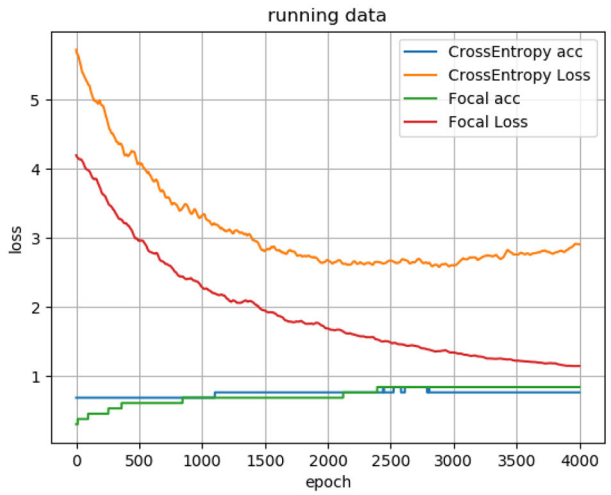

FIGURE 5 | Loss curve and accuracy curve of using focal loss function and cross-entropy loss function for the 1D CNN.

setting $r > 0$ can reduce the loss weight of easy-to-predict samples, which can make the model pay more attention to the difficult-to-predict samples during training. Through many experiments, we set α to 0.2 and r to 2. Moreover, the difference between using the focal loss function and the ordinary cross-entropy loss function for the neural network is shown in **Figure 5**.

However, ELM does not need to adjust the parameters by iteratively reducing the loss. When the input weight and the bias of the hidden layer are randomly determined, the output matrix of the hidden layer is uniquely determined. The training of the neural network is transformed into solving a linear system:

$$\mathbf{H}\beta = \mathbf{T}, \tag{4}$$

where H is the output of the hidden layer node; β is the output weight, and T is the expected output. We can obtain the output weight β by transforming H into the generalized inverse matrix $H^{'}$ and multiplying T.

At present, LSTM is the most popular model in processing time series, which can solve the problem of long-term dependence on information very well. So, this paper also takes this model into account and compares it with 1D CNN. The LSTM model used in this paper consists of 32 basic units.

Evaluation Indicators

There are many indicators to evaluate the quality of a neural network. However, considering that this research involves the auxiliary diagnosis of diseases, this article used three evaluation indicators, including accuracy (ACC), sensitivity (SES), and specificity (SPC), which were expressed as follows:

$$\mathbf{ACC} = \frac{\mathbf{TP} + \mathbf{TN}}{\mathbf{TP} + \mathbf{FP} + \mathbf{FN} + \mathbf{TN}}, \tag{5}$$

$$\mathbf{SES} = \frac{\mathbf{TP}}{\mathbf{TP} + \mathbf{FN}}, \tag{6}$$

$$\mathbf{SPC} = \frac{\mathbf{TN}}{\mathbf{TN} + \mathbf{FP}}, \tag{7}$$

where TP, TN, FN, and FP indicate true positive, true negative, false negative, and false positive, respectively.

In this paper, Keras was used as a deep learning model framework, and TensorFlow was selected as the backend, which created a 1D CNN model. Meanwhile, the experimental environment was CUDA 10.1; the GPU was NVIDIA GeForce GTX 1080; the CPU was Intel Core i7-8700, and the operating system was Windows 10.

RESULTS

We tested each model on three different data sets of the subjects, including walking data, running data, and the combination of walking and running data to explore the pros and cons of the models as a whole. The three data sets were divided similarly, and 70% of the data sets were randomly selected for training, and the training data were subjected to 10-fold cross-validation to obtain the optimal model parameters. Then, the remaining 30% of the data were used for testing. Considering that our data set was small, the batch size of the network was set to the entire training set. Using this method, the loss direction determined by the full data set could represent the sample population, thereby moving accurately toward the direction of the extreme value.

We repeated each experiment 10 times independently and took the average of the results as the judgement of the model. For the data division of each trial, the data distribution in the training set and test set was the same.

Comparison Results of all Neural Network Models

The overall results are shown in **Tables 2–4**. It can be seen from the figure that all the neural network models have the best effect on the running data set. In order to make the comparison results on the running data set more visible, this paper makes a histogram, as shown in **Figure 6**.

Results of Attention Mechanism

According to the comparison results in the previous section, this paper will make further research on the running data set. The soft attention mechanism could reweight all information adaptively before aggregation. Consequently, important information could be separated, and the interference of unimportant information could be avoided to improve the accuracy. In this study, the weight of time dimension was fixed, and only the input feature dimension was weighted. After the neural network model was trained, the weight of feature dimension was determined. Finally, we visualized the weight assigned to each feature by the attention

TABLE 2 | Results on walking data set.

Network models	ACC	SES	SPC	Training time(s)
1D CNN	0.68	0.77	0.53	43.8
2D CNN	0.61	0.81	0.29	158
LSTM	0.63	0.73	0.51	153.1
VGGNet	0.61	0.91	0.14	1913
AlexNet	0.61	1.00	0.00	800.7
ELM	0.66	0.89	0.29	0.02
BP	0.58	0.59	0.55	4.32

TABLE 3 | Results on running data set.

Network models	ACC	SES	SPC	Training time(s)
1D CNN	0.924	0.97	0.84	43.7
2D CNN	0.64	0.81	0.35	157.4
LSTM	0.79	0.83	0.69	152
VGGNet	0.74	0.80	0.59	1912.4
AlexNet	0.769	0.88	0.60	800.5
ELM	0.71	0.88	0.51	0.03
BP	0.65	0.87	0.32	4.42

TABLE 4 | Results on combined walking and running data set.

Network models	ACC	SES	SPC	Training time(s)
1D CNN	0.77	0.77	0.70	43.8
2D CNN	0.615	0.88	0.20	160
LSTM	0.76	0.90	0.58	155
VGGNet	0.62	0.80	0.42	1914
AlexNet	0.76	0.84	0.64	801
ELM	0.59	0.82	0.20	0.03
BP	0.56	0.66	0.40	4.51

mechanism and observed the features that belonged to the key features (**Figure 7**).

Visualization Results of the CNN Model

In this section, the T-SNE method was used to visualize the feature distribution of the input layer, final convolution layer, and output layer of the four CNN models for running data set. In this way, we can easily compare the ability of learning features from the original biomechanical data among different CNN models **Figure 8**).

DISCUSSION

As shown in **Tables 2–4**, all the neural network models perform best in the running data set, which indicates that PFPS will have a significant impact on the lower limb biomechanical features of patients during running. Pain is a protective mechanism for patients, and patients will take corresponding compensatory behavior to complete the exercise to reduce pain, thereby resulting in changes in biomechanical features. The task intensity of running is higher than that of walking, which may lead to evident compensatory changes in patients with pain, thereby making the neural network easier to learn.

By adding attention mechanism into the 1D CNN model and outputting the weight results of attention mechanism, we ranked the importance of biomechanical features in identifying PFPS and determined the biomechanical features that were important for the identification of PFPS. As shown in **Figure 6**, the three most concerned features of the neural network are VM, SEB, and KF. However, whether the changes of these biomechanical features cause PFPS, or whether the pain of PFPS causes the changes of these biomechanical features, that is, whether these biomechanical features are risk factors for PFPS, remain unclear.

All neural network models have high specificity and low sensitivity. There are two reasons for this result. First, more PFPS patients are included in the data set, which makes the learning of the network prone to deviation. Second, the data set is relatively small, which makes the neural network easy to overfit. Previous

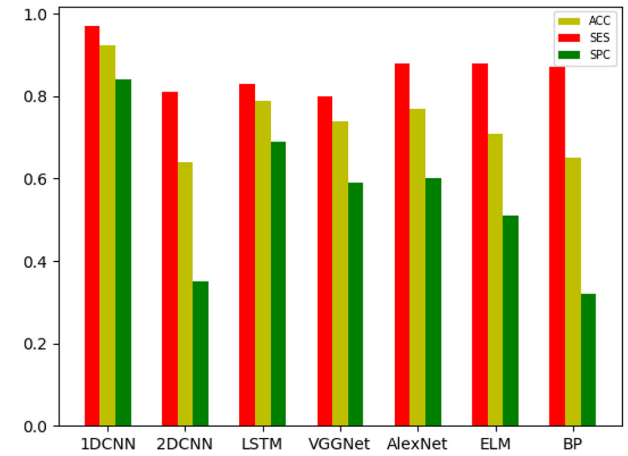

FIGURE 6 | The results of each neural network on the running data set.

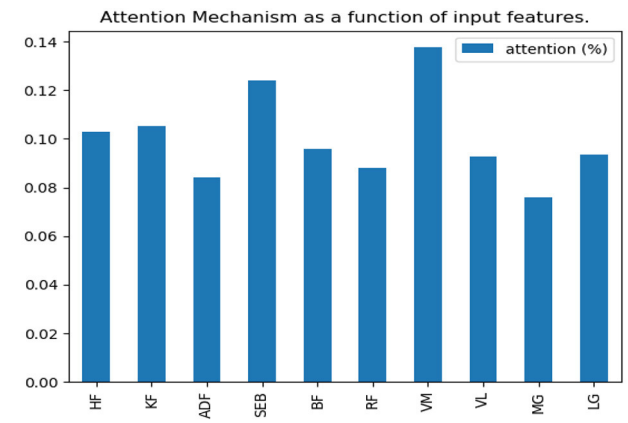

FIGURE 7 | Attention probability distribution of input features on running data set.

studies have shown that CNN tends to perform better in big data. In the case of a larger data set, we hypothesize that the accuracy of our model can be improved. In addition, although ELM and BP are feedforward neural networks with a hidden layer, the number of hidden layer nodes is different when they reach the optimal situation probably because their network weights are determined in different ways. ELM directly determines the weight of neurons in the hidden layer by solving the generalized inverse matrix, whereas BP gradually determines the weight of neurons in the hidden layer by back propagation.

In all data sets, the 1D CNN performs best, particularly on the running data set (ACC = 0.924, SES = 0.97, SPC = 0.84). Meanwhile, the comparison of the classification results shows that the 1D CNN is suitable for the characteristics of these biomechanical parameters than the 2D CNN. In addition, the introduction of focal loss does not greatly improve the accuracy of the neural network, but it makes the neural network easier to learn to ensure that the SES and SPC values will not differ remarkably. The results of 1D CNN are also better than

FIGURE 8 | Visualization of feature representations extracted from input layer, last convolutional layer and output layer for running data set.

LSTM, which may be because 1D CNN pays more attention to the feature changes in local time period, while LSTM is more suitable for data with long-term dependence. The disease detection of pain type should pay more attention to the instant changes caused by pain. Moreover, because LSTM adopts full connection computing mode, its computation is very time-consuming, resulting in poor real-time performance. Compared with the LSTM model, the local connection and weight sharing mechanism of the 1D CNN model reduces a large number of network parameters, so that the model can train faster and reduce the risk of overfitting.

In this paper, the *t*-SNE method was used to reduce the dimension and visualize the features extracted from the CNN model and determine whether the features extracted from the neural network model were separable, which increased the interpretability of the model. As shown in **Figure 7**, the 1D CNN model constructed in this paper could easily obtain segment able features.

CONCLUSION

This paper proposed a method to assist the diagnosis of PFPS through the 1D CNN model. Different from previous studies, this method does not require complex biomechanical models, and it can achieve high accuracy (ACC = 0.924) only through some directly measurable biomechanical parameters and the gender of subjects. This method is easy to operate. After the neural network has learned a certain number of features, the model is saved. Then, the PFPS can be intelligently determined by the neural network in real-time through the lower limb

joint angle values and sEMG signals of subjects in a gait cycle. This prospective study provides new insights into the auxiliary diagnosis of PFPS, which can be used to develop a convenient, efficient, and universal auxiliary diagnosis model for PFPS.

Compared with previous research (2, 37), the method of this study has higher sensitivity (SES = 97%), and specificity (SPC = 84%). Ferrari et al. used the mid-band parameters of sEMG, which were associated with anterior knee pain to determine PFPS. The method had 70% sensitivity and 87% specificity, and the trial involved 51 subjects, including 22 PFPS patients and 29 painless controls (2). Briani et al. used the sEMG signal of VM to diagnose PFPS, and obtained 72% sensitivity and 69% specificity, and obtained 68% sensitivity and 62% specificity through the sEMG signal of VL. The trial involved 59 subjects, including 31 patients with PFPS and 28 painless controls (37).

This study is a preliminary investigation, and its applicability requires caution. This study has some limitations, which need to be considered in future studies. For example, a comparative experiment should be conducted to explore whether these biomechanical changes caused by pain or PFPS caused by these biomechanical changes. Another limitation is that the data set of the paper is relatively small, and the convolutional neural network often performs better on large data sets; therefore, larger sample size must also be considered in the next work. Meanwhile, future work must focus on the specific subclassifications of PFP diagnoses.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

WS, BX, and MD conceived the layout, rationale, and plan of this manuscript. WS wrote the first draft of the manuscript. CL, JL, YZ, QZ, DX, and YL edited the manuscript. All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was supported by Natural Science Foundation of Fujian Province (2020J01890).

ACKNOWLEDGMENTS

The authors would like to thank all the participants.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2021.615597/full#supplementary-material

REFERENCES

1. Fagan V, Delahunt E. Patellofemoral pain syndrome: a review on the associated neuromuscular deficits and current treatment options. *Br J Sports Med.* (2008) 42:489–95. doi: 10.1136/bjsm.2008.046623
2. Ferrari D, Kuriki HU, Silva CR, Alves N, Mícolis de Azevedo F. Diagnostic accuracy of the electromyography parameters associated with anterior knee pain in the diagnosis of patellofemoral pain syndrome. *Arch Phys Med Rehabil.* (2014) 95:1521–6. doi: 10.1016/j.apmr.2014.03.028
3. Tuna BK, Semiz-Oysu A, Pekar B, Bukte Y, Hayirlioglu A. The association of patellofemoral joint morphology with chondromalacia patella: a quantitative MRI analysis. *Clin Imaging.* (2014) 38:495–8. doi: 10.1016/j.clinimag.2014.01.012
4. Dehaven KE, Lintner DM. Athletic injuries: comparison by age, sport, and gender. *Am J Sports Med.* (1986) 14:218–24. doi: 10.1177/036354658601400307
5. Barton CJ, Lack S, Hemmings S, Tufail S, Morrissey D. The 'best practice guide to conservative management of patellofemoral pain': incorporating level 1 evidence with expert clinical reasoning. *Br J Sports Med.* (2015) 49:923–34. doi: 10.1136/bjsports-2014-093637
6. Powers CM, Bolgla LA, Callaghan MJ, Collins N, Sheehan FT. Patellofemoral Pain: Proximal, Distal, and Local Factors-—2nd International Research Retreat, August 31–September 2, 2011, Ghent, Belgium. *J Orthop Sports Phys Ther.* (2012) 42:A1–54. doi: 10.2519/jospt.2012.0301
7. Thomas MJ, Wood L, Selfe J, Peat G. Anterior knee pain in younger adults as a precursor to subsequent patellofemoral osteoarthritis: a systematic review. *BMC Musculoskelet Disord.* (2010) 11:201. doi: 10.1186/1471-2474-11-201
8. Cook C, Mabry L, Reiman MP, Hegedus EJ. Best tests/clinical findings for screening and diagnosis of patellofemoral pain syndrome: a systematic review. *Physiotherapy.* (2012) 98:93–100. doi: 10.1016/j.physio.2011.09.001
9. Dye SF. The pathophysiology of patellofemoral pain: a tissue homeostasis perspective. *Clin Orthop Relat Res.* (2005) 436:100–10. doi: 10.1097/01.blo.0000172303.74414.7d
10. Jensen R, Kvale A, Baerheim A. Is pain in patellofemoral pain syndrome neuropathic? *Clin J Pain.* (2008) 24:384–94. doi: 10.1097/AJP.0b013e3181658170
11. Csintalan RP, Schulz MM, Woo J, Mcmahon PJ, Lee TQ. Gender differences in patellofemoral joint biomechanics. *Clin Orthop Relat Res.* (2002) 402:260–9. doi: 10.1097/00003086-200209000-00026
12. Nunes GS, Stapait EL, Kirsten MH, De Noronha M, Santos GM. Clinical test for diagnosis of patellofemoral pain syndrome: Systematic review with meta-analysis. *Phys Ther Sport.* (2013) 14:54–9. doi: 10.1016/j.ptsp.2012.11.003
13. Mustamsir E, Phatama KY, Pratianto A, Pradana AS, Hidayat M. Validity and reliability of the indonesian version of the kujala score for patients with patellofemoral pain syndrome. *Orthop J Sports Med.* (2020) 8:232596712092294. doi: 10.1177/2325967120922943
14. Chen S, Chang W-D, Wu J-Y, Fong Y-C. Electromyographic analysis of hip and knee muscles during specific exercise movements in females with patellofemoral pain syndrome: an observational study. *Medicine.* (2018) 97:e11424. doi: 10.1097/MD.0000000000011424
15. Dag F, Dal U, Altinkaya Z, Erdogan AT, Ozdemir E, Yildirim DD, et al. Alterations in energy consumption and plantar pressure distribution during walking in young adults with patellofemoral pain syndrome. *Acta Orthopaedica et Traumatologica Turcica.* (2019) 53:50–5. doi: 10.1016/j.aott.2018.10.006
16. Liew BXW, Abichandani D, De Nunzio AM. Individuals with patellofemoral pain syndrome have altered inter-leg force coordination. *Gait Posture.* (2020) 79:65–70. doi: 10.1016/j.gaitpost.2020.04.006

17. Besier TF, Fredericson M, Gold GE, Beaupré GS, Delp SL. Knee muscle forces during walking and running in patellofemoral pain patients and pain-free controls. *J Biomech.* (2009) 42:898–905. doi: 10.1016/j.jbiomech.2009.01.032
18. Myer GD, Ford KR, Foss KDB, Rauh MJ, Hewett TE. A predictive model to estimate knee-abduction moment: implications for development of a clinically applicable patellofemoral pain screening tool in female athletes. *J Athl Train.* (2014) 49:389–98. doi: 10.4085/1062-6050-49.2.17
19. Zeng N, Wang Z, Zhang H. Inferring nonlinear lateral flow immunoassay state-space models via an unscented Kalman filter. *Sci China Inf Sci.* (2016) 59:112204. doi: 10.1007/s11432-016-0280-9
20. Schöllhorn WI. Applications of artificial neural nets in clinical biomechanics. *Clin Biomech.* (2004) 19:876–98. doi: 10.1016/j.clinbiomech.2004.04.005
21. Zeng N, Qiu H, Wang Z, Liu W, Zhang H, Li Y. A new switching-delayed-PSO-based optimized SVM algorithm for diagnosis of Alzheimer's disease. *Neurocomputing.* (2018) 320:195–202. doi: 10.1016/j.neucom.2018.09.001
22. Xiong B, Zeng N, Li Y, Du M, Huang M, Shi W, et al. Determining the online measurable input variables in human joint moment intelligent prediction based on the hill muscle model. *Sensors.* (2020) 20:1185. doi: 10.3390/s20041185
23. Keijsers NLW, Stolwijk NM, Louwerens JWK, Duysens J. Classification of forefoot pain based on plantar pressure measurements. *Clin Biomech.* (2013) 28:350–6. doi: 10.1016/j.clinbiomech.2013.01.012
24. Otag I, Otag A, Akkoyun S, Çimen M. A way in determination of patellar position: ligamentum patellae angle and a neural network application. *Biocybern Biomed Eng.* (2014) 34:184–8. doi: 10.1016/j.bbe.2014.02.004
25. Lee S, Kim H, Lieu QX, Lee J. CNN-based image recognition for topology optimization. *Knowl Based Syst.* (2020) 198:105887. doi: 10.1016/j.knosys.2020.105887
26. Huynh HX, Nguyen VT, Duong-Trung N, Pham VH, Phan CT. Distributed framework for automating opinion discretization from text corpora on facebook. *IEEE Access.* (2019) 7:78675–84. doi: 10.1109/ACCESS.2019.2922427
27. Mishra SR, Mishra TK, Sanyal G, Sarkar A, Satapathy SC. Real time human action recognition using triggered frame extraction and a typical CNN heuristic. *Pattern Recognit Lett.* (2020) 135:329–36. doi: 10.1016/j.patrec.2020.04.031
28. Wang S, Sun J, Mehmood I, Pan C, Chen Y, Zhang Y-D. Cerebral micro-bleeding identification based on a nine-layer convolutional neural network with stochastic pooling. Concurr Comput. (2020) 32:e5130. doi: 10.1002/cpe.5130
29. Wang S-H, Govindaraj VV, Górriz JM, Zhang X, Zhang Y-D. Covid-19 classification by FGCNet with deep feature fusion from graph convolutional network and convolutional neural network. *Inf Fusion.* (2021) 67:208–29. doi: 10.1016/j.inffus.2020.10.004
30. Wang S-H, Zhang Y-D. DenseNet-201-based deep neural network with composite learning factor and precomputation for multiple sclerosis classification. *ACM Trans Multimedia Comput Commun Appl.* (2020) 16:1–19. doi: 10.1145/3341095
31. Krizhevsky A, Sutskever I, Hinton GE. *ImageNet Classification with Deep Convolutional Neural Networks*. NIPS, Vol. 25. Curran Associates Inc. (2012).
32. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv.* (2014).
33. Lu S, Wang S-H, Zhang Y-D. Detection of abnormal brain in MRI via improved AlexNet and ELM optimized by chaotic bat algorithm. *Neural Comput Appl.* (2020). doi: 10.1007/s00521-020-05082-4. [Epub ahead of print].
34. Lu S, Lu Z, Zhang Y-D. Pathological brain detection based on AlexNet and transfer learning. *J Comput Sci.* (2019) 30:41–7. doi: 10.1016/j.jocs.2018.11.008
35. Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. *IEEE Trans Pattern Anal Mach Intell.* (2017) 42:2999–3007. doi: 10.1109/TPAMI.2018.2858826
36. Romdhane TF, Alhichri H, Ouni R, Atri M. Electrocardiogram heartbeat classification based on a deep convolutional neural network and focal loss. *Comput Biol Med.* (2020) 123:103866. doi: 10.1016/j.compbiomed.2020.103866
37. Briani RV, de Oliveira Silva D, Pazzinatto MF, de Albuquerque CE, Ferrari D, Aragão FA, et al. Comparison of frequency and time domain electromyography parameters in women with patellofemoral pain. *Clin Biomech.* 2015 30:302–7. doi: 10.1016/j.clinbiomech.2014.12.014

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Artificial Intelligence

ORIGINAL RESEARCH
published: 04 June 2021
doi: 10.3389/frai.2021.642731

Evaluation of MRI Denoising Methods Using Unsupervised Learning

Marc Moreno López[1]*, *Joshua M. Frederick*[2] and *Jonathan Ventura*[2]

[1]Department of Computer Science, University of Colorado Colorado Springs, Colorado Springs, CO, United States, [2]Department of Computer Science and Software Engineering, California Polytechnic State University, San Luis Obispo, CA, United States

In this paper we evaluate two unsupervised approaches to denoise Magnetic Resonance Images (MRI) in the complex image space using the raw information that k-space holds. The first method is based on Stein's Unbiased Risk Estimator, while the second approach is based on a blindspot network, which limits the network's receptive field. Both methods are tested on two different datasets, one containing real knee MRI and the other consists of synthetic brain MRI. These datasets contain information about the complex image space which will be used for denoising purposes. Both networks are compared against a state-of-the-art algorithm, Non-Local Means (NLM) using quantitative and qualitative measures. For most given metrics and qualitative measures, both networks outperformed NLM, and they prove to be reliable denoising methods.

Keywords: deep learning, denoising, k-Space, MRI, unsupervised

OPEN ACCESS

Edited by:
Shuai Li,
Swansea University, United Kingdom

Reviewed by:
Shivanand Sharanappa Gornale,
Rani Channamma University, India
Ameer Tamoor Khan,
Hong Kong Polytechnic University, China

***Correspondence:**
Marc Moreno López
mmorenol@uccs.edu

Specialty section:
This article was submitted to Medicine and Public Health, a section of the journal Frontiers in Artificial Intelligence

Received: *16 December 2020*
Accepted: *17 May 2021*
Published: *04 June 2021*

Citation:
Moreno López M, Frederick JM and Ventura J (2021) Evaluation of MRI Denoising Methods Using Unsupervised Learning. Front. Artif. Intell. 4:642731. doi: 10.3389/frai.2021.642731

1 INTRODUCTION

Magnetic Resonance Imaging, MRI, is one of the most widely used imaging techniques, as it provides detailed information about organs and tissues in a completely non-invasive way. In MRI, data needed to generate images is directly sampled from the spatial frequency domain; however, the quality of this data can be deteriorated by several thermal noise sources and artifacts. Noise in MRI is of major consequence as it can mislead and result in inaccurate diagnoses of patients. In addition to visually corrupting the recovered images, noise is also an obstacle when conducting quantitative imaging on the MRI. The utility of MRI decreases if a region or specific tissue suffers from a low signal to noise ratio. Thus, there is a necessity for an efficient MRI reconstruction process, where denoising methods are applied to noisy images in order to improve both qualitative and quantitative measures of MRI.

Additionally, in the case of *in vivo* MRI, noise is implicit to the acquisition process. When taking an MRI of a living subject, there are multiple noise factors. All other factors withheld, the MR machine has an innate noise component when acquiring an image due to a thermal factor. Another source of thermal noise is inversely proportional to the amount of time that the subject stays inside the MR machine, and while in the machine the subjects movements also contribute to the thermal noise. Finally, the patient's body temperature and the thermal factor from the MR machine is another key element, specially since a long exposure inside the MR machine could lead to an increase in body temperature, web (2017).

Thus, when training a MRI denoiser, no ground truth is available for the training procedure. Likewise, due to previously discussed movement of the subject, two independent samples for denoising strategies as used by Lehtinen et al. (2018) cannot be reasonably obtained. Thus either synthetic data needs be generated for supervised learning or unsupervised and self-supervised strategies must be employed. As such, we evaluate self-supervised solutions to MRI denoising. Deep self-supervised image denoisers have been seeing recent success for general image denoising tasks, and provide robust denoisers without requiring access to denoised images. Self-supervised denoisers

generally under-perform supervised techniques, but arise naturally in cases like MRI, where pure supervised learning is infeasible.

While deep learning has seen success in many areas, there is a lack of methods focused on denoising MRI. Additionally, many traditional techniques denoise MRI in the magnitude space, dismissing the innate spatial frequency information the MRI contain. Most of the MRI denoising methods available use a supervised approach where they use the original MRI as ground truth. We wanted to explore an unsupervised approach using the complex image space, where no ground truth data is needed. Therefore, we will compare two unsupervised denoising approaches that denoise MRI in the spatial frequency space, competing with the more classical and widely used denoising methods.

2 MATERIALS AND METHODS

2.1 Related Work

Previous attempts on MRI denoising can be categorized in three different ways: traditional methods, supervised learning, and unsupervised learning.

2.1.1 Traditional Methods

Traditional MRI denoising techniques are generally based on filtering, transformations, or statistical methods such as Mohan et al. (2014). Three of the most widely-used methods currently are bilateral filtering by Tomasi and Manduchi (1998), non-local means by Buades et al. (2005), and BM3D by Dabov et al. (2007).

The bilateral filter presented by Tomasi and Manduchi (1998) is an edge preserving non-iterative method. When applied to an image, it uses a low-pass denoising kernel which adjusts to the original image spatial distribution of pixel-values. This helps preserve the edges while denoising the image. In the presence of sharp transitions, the kernel is weighted according to this transition. This behavior is modeled by a convolution of the intensity values of the image and a non-linear weighting function.

Non-local means, Buades et al. (2005), or NLM, uses the self spatial similarities that natural images have. It exploits the redundancy of the neighborhood pixels to remove the noise. The simplicity of this filter consists of using those similarities to find similar patches on the rest of the image to the patch being denoised. This is known as neighborhood filtering. NLM assigns confidence weights based on similarity to the original patch and its distance from the center of the observed patch. The main issue with NLM is that since it relies on a large space search, it can create a bottleneck in terms of computation.

BM3D, Dabov et al. (2007), is a robust algorithm that has several parameters that can be modified in order to achieve the best denoising. It is an extension of NLM, in the sense that it uses spatial similarities within the image. It starts by searching for patches with similar intensities to the patch that is being denoised. A 3D matrix containing the size of the patch and the aggregated patches is built. Then, a 3D transform is applied. So as to remove high frequency noises, the transform space is filtered and thresholded. Finally, a denoised 3D block is yielded by doing the inverse transformation. To recover the original array, weights are assigned to every patch. These weights are based on the variance and distance of the patch.

2.1.2 Supervised Learning

One of the most well-known approaches for supervised denoising, DnCNN, is presented by Zhang et al. (2017). Their method uses feed-forward Convolutional Neural Networks, CNN. In order to improve both algorithm speed and performance, they use residual modules and batch normalization. This makes their network unique. Also, it does not need to know the level of noise. So, it can perform blind Gaussian denoising.

Bermudez et al. (2018) implemented an autoencoder with skip connections. To test their method, they added Gaussian noise to a T1-weighted brain MRI dataset from healthy subjects. Benou et al. (2017) worked on spatio-temporal denoising of brain MRI using ensembles of deep neural networks. Each network is trained on a different variations of SNR. By doing this, they generate different hypothesis and then select the most likely one to generate a clean output curve using a classification network. This method presented better denoising results than those presented by Gal et al. (2010), where they use a dynamic NLM method, and they were also better than the results presented by Vincent et al. (2010), where they use stacked denoising autoencoders. An interesting approach is presented by Jiang et al. (2018). They use a multi-channel DnCNN to denoise Rician noise in magnitude MRI instead of Gaussian noise. They test their network for both known and unknown levels of noise, which allows them to create a more general model. Finally, Tripathi and Bag (2020) present a CNN with residual learning to denoise synthetic brain MRI. They use five different clean synthetic magnitude datasets and add Rician noise to it. They also perform blind denoising, where the network is tested with a different level of noise than it was trained with. Their blind denoising test yields interesting results, since they prove that, when the network is trained with higher levels of noise and tested on lower levels of noise, the network yields better results than when training and testing with low noise.

2.1.3 Unsupervised Learning

For unsupervised image denoising a novel method is presented by Xu et al. (2020), where they introduce a method that uses corrupted test images as their ground truth "clean" images. To train their network they use synthetic images consisting of small alterations to the corrupted test image. They add more noise to the test image, and they prove that if they introduce a small amount of noise to the test image as an alteration, their network is still capable of denoising the corrupt image and produce a clean output. Given their training methodology, which trains an image-specific network for each image to be denoised, their approach is not well suited for MRI denoising, given the volume of images contained in an MRI. Therefore, the denoising process would be too time-consuming.

One of the most effective models used for unsupervised denoising is presented by Soltanayev and Chun (2018) and it

is based on Stein's unbiased risk estimator, SURE. The SURE estimator, presented by Stein (1981) is an unbiased MSE estimator. The only problem with the SURE estimator is that it can only be expressed in an analytical form. When this is not available, Ramani et al. (2008) proposed a Monte-Carlo-based SURE, MC-SURE. The work presented by Soltanayev and Chun (2018) overcomes previous shortcomings and combines the Monte-Carlo approximation and makes it available for deep neural network models. Since it can be used with no need of noiseless ground truth data, deep neural networks can be trained for denoising purposes in an unsupervised manner.

The model Noise2Noise (N2N) by Lehtinen et al. (2018), saw success in denoising images by learning to predict one noisy image from another by training on independent pairs of noisy images. The result is a model that predicts the expected value of the noisy distribution for each pixel. For many real noise models, Gaussian, Poisson, etc, this expected value is clean signal.

Building upon this, Noise2Void (N2V) by Krull et al. (2018) developed a strategy which removes the need for two independent samples, and instead learns to denoise an image in a fully self-supervised way. In place of a second independent sample, N2V learns to denoise from the receptive field of a single pixel, excluding itself.

Using this strategy, Noise2Self developed a general framework for this type of denoising problem for higher dimensional spaces, and Laine et al. (2019) denoted this form of network as a "blindspot" network and provide several improvements.

Despite all the progress in unsupervised denoising in other areas, there is not that much work done in unsupervised MRI denoising. One example is by Eun et al. (2020), where they introduce a cycle generative adversarial network, CycleGAN to denoise compressed sensing MRI. Thus, we wanted to further explore this path, given the potential that unsupervised learning showed in other fields and the lack of clean ground truth data when working with MRI.

2.2 Background

2.2.1 K-Space

In MRI terminology, k-space is the 2D or 3D Fourier transform of the MRI measured. When measuring an MRI, the complex values are sampled using a pulse sequence, such as radio-frequency and gradient pulses. At the end of the scan the data is mathematically processed to produce a final image. Therefore k-space holds raw data before reconstruction. K-space can be seen as an array of numbers representing spatial frequencies in the MRI.

To transition between k-space and the complex image space, we apply an inverse fast Fourier transform, and vice versa. Even though they are visually different, the information contained in both spaces is the exactly the same. In k-space, the axes represent spatial frequencies instead of positions. The points plotted in this space do not correspond one on one to the pixels on the image in time domain. Every point in k-space contains information about phase and spatial frequency for every pixel in the time as seen in **Figure 1**.

In MRI, the thermal noise that deteriorates the k-space is Gaussian. This Gaussian noise model can be defined as $y = x + n$, where x is the original MRI signal and n is Gaussian noise. Even after applying the inverse fast Fourier transform, the noise remains Gaussian. If we converted the complex MRI to magnitude MRI, then the noise would be Rician. This is why, we want to explore Gaussian denoising of complex-value data and avoid dealing with Rician noise in the magnitude space.

2.2.2 SURE Estimator

When training a network, a gradient-based optimization algorithm is used such as the stochastic gradient descent (SGD) Bottou (1999), momentum, or the Adam optimization algorithm Kingma and Ba (2015) to optimize the loss. In our case, we use the Mean Squared Error, MSE web (2020a), to calculate the amount of noise present in the image.

$$\frac{1}{M}\sum_{j=1}^{M}\left\|h\left(y^{(j)};\theta\right)-x^{(j)}\right\|^2 \tag{1}$$

where M is the number of samples in one batch of data. The main issue with **Eq. 1** is that, since we are working in an unsupervised environment, we do not have access to x, the ground truth. Therefore, an estimator for MSE needs to be used. This is done by the SURE estimator presented in **Eq. 2**

$$\frac{1}{M}\sum_{j=1}^{M}\left[\left\|\boldsymbol{y}^{(j)}-\boldsymbol{h}\left(\boldsymbol{y}^{(j)};\boldsymbol{\theta}\right)\right\|^2-K\sigma^2+2\sigma^2\sum_{i=1}^{K}\frac{\partial \boldsymbol{h}_i\left(\boldsymbol{y}^{(j)};\boldsymbol{\theta}\right)}{\partial \boldsymbol{y}_i}\right] \tag{2}$$

noting that no noiseless ground truth data were used in **Eq. 2**.

The only problem with the SURE estimator is that the last divergence is intractable. However it can be approximated using the Monte-Carlo SURE estimator by Ramani et al. (2008). Therefore the final risk estimator which will be used as a loss function is

$$\frac{1}{M}\sum_{j=1}^{M}\left\{\boldsymbol{y}^{(j)}-\boldsymbol{h}\left(\boldsymbol{y}^{(j)};\boldsymbol{\theta}\right)^2-K\sigma^2+\frac{2\sigma^2}{\epsilon}\left(\tilde{\boldsymbol{n}}^{(j)}\right)^{\mathrm{t}}\left[\boldsymbol{h}\left(\boldsymbol{y}^{(j)}+\epsilon\tilde{\boldsymbol{n}}^{(j)};\boldsymbol{\theta}\right)-\boldsymbol{h}\left(\boldsymbol{y}^{(j)};\boldsymbol{\theta}\right)\right]\right\} \tag{3}$$

where ε is a small fixed positive number and $n \sim (j)$ is a single realization from the standard normal distribution for each training data j.

2.2.3 Blindspot Network

Laine et al. (2019) provide an improved blindspot architecture and denoising procedure. The blindspot network architecture combines multiple branches, where each branch restricts its receptive field to a half-plane which does not contain the center pixel. Then four branches are combined using 1×1 convolutions. This form allows for the receptive field to be efficiently extended arbitrarily in every direction, while still excluding the center pixel.

In N2V, the center pixel information is not exploited to prevent the model from simply learning to output this value. However, using Bayesian reason to the denoising task, we have for a particular noisy pixel y and corresponding clean signal x

$$p(x \mid y, \Omega_y) \propto p(y \mid x) p(x \mid \Omega_y) \tag{4}$$

where Ω_y is the context given by the receptive field of the pixel y. Thus, using a blindspot architecture to model a Gaussian prior $p(x \mid \Omega_y)$, the posterior mean $\mathcal{E}_x[p(x \mid y, \Omega)]$ has a closed form solution for many noise models. This allows for the use of the previously unexploited center pixel data at test time. In the case of MRI, with a Gaussian noise model, the posterior mean can be computed analytically.

2.2.4 Datasets

2.2.4.1 Knee MRI

The Center for Advanced Imaging Innovation and Research (CAI^2R), in the Department of Radiology at New York University, NYU, School of Medicine and NYU Langone Health, released two MRI datasets, Zbontar et al. (2018), Zbontar et al. (2020), to work on rapid image acquisition and advanced image reconstruction. The deidentified datasets consist of scans of knees and brains, which contain raw k-space data. For this experiment, we decided to use single coil data only, as it is the most widely used modality and due to its data size compared to multi coil, which is smaller.

The knee single coil dataset contains 973 training subjects and 199 validation subjects. According to their website, the fully sampled knee MRIs were obtained on 3 and 1.5 Tesla magnets. The raw dataset includes coronal proton density-weighted images with and without fat suppression. As such, NYU fastMRI investigators provided data but did not participate in analysis or writing of this report. A listing of NYU fastMRI investigators, subject to updates, can be found at web (2020b).

Note that all knee MRI contain noise that varies from subject to subject.

2.2.4.2 Brainweb

In most of today's image analysis methods, a ground truth is expected, even if just for validation. In the case of MRI, noise is implicit to the *in vivo* acquisition process, and so no true noise free MR dataset exists. The Brainweb dataset provides an easy solution for this by creating a Simulated Brain Database (SBD) Cocosco et al. (1997); web (1998); Kwan et al. (1999); Kwan et al. (1996); Collins et al. (1998), where an MRI simulator is used to created realistic MRI data volumes. In addition to providing a predefined magnitude image dataset, the Brainweb simulator is exposed to allow for custom simulations.

Using the custom simulator, we acquired raw frequency spatial data for varied simulator parameters. This includes data generated for all combinations of no, mild, moderate, and severe multiple sclerosis (MS) lesions anatomic models with the six available parameter templates. These six are generated by combining the AI and ICBM protocols with either T1, T2, or Proton Density (PD) weighting. For our purposes, we will only be using T1 and T2. All together this allowed for the generation of 16 brain MR volumes simulated from a realistic parameter set. 12 subjects were used for training and four subjects were used for testing. Additionally, the custom simulator allows for adding a noise level; however, as we are treating this data as ground truth, we did not use this feature. For all Brainweb experiments, we performed cross-validation to ensure the validity of the results.

Since our blindspot network expects square input, each individual slice of the MR volumes were zero padded in k-space to have matching dimensions.

2.3 Training

All models were trained and tested using a single NVIDIA GeForce GTX Titan X, with 12 GBytes of memory.

2.3.1 SURE Model

The gradient of **Eq. 3** can be automatically calculated when training a deep learning framework. Therefore, we use **Eq. 3** as a cost function for a basic U-Net architecture, Ronneberger et al. (2015), with five convolutional layers on both sides.

To train the SURE estimator in 2D, we use a U-Net of depth 5, convolution kernel size of 3 and 48 initial feature maps. After each convolutional layer, a LeakyReLU is applied, except for the last convolutional layer, where no activation function is used. We train the network in batches of 10 for 300 epochs, using the Adam optimizer with an initial learning rate of 3×10^{-4}. The data, both training and testing, is center cropped to 320 × 320 for knee MRI and 192 × 192 for brain MRI, using all available slices for both.

2.3.2 Blindspot Model

Due to large regions of no-signal in MRI and a shared standard deviation across all pixels, many techniques exist to estimate the standard deviation of the noise σ, Sardy et al. (2001). Thus, we use a blindspot architecture with knowledge of σ, and our prior becomes $p(x \mid \Omega_y, \sigma)$. This modifies **Eq. 4** in training to

$$p(x \mid y, \Omega_y) \propto p(y \mid x)\, p(x \mid \Omega_y, \sigma)$$

We train a 5-layer deep blindspot network in batches of 5 for 300 epochs. The convolution kernel has size of 3 and there are 48 initial feature maps. No activation function is used. We use Adam optimizer with an initial learning rate of 3×10^{-4}. The learning rate is reduced if the validation loss has not decreased after ten epochs. The data, both training and testing, is center cropped to 320 × 320 for knee MRI and 192 × 192 for brain MRI, using all available slices for both. For a more detailed network architecture description, please refer to Laine et al. (2019). We used the same blindspot network and U-Net architecture as described in Laine et al. (2019).

3 RESULTS

For both datasets, different levels of noise were added to the original images in order to do a quantitative comparison to NLM. Since both models rely on Gaussian noise, we will only be adding Gaussian noise to the images.

For the knee single coil dataset, we started by adding noise with $\sigma = 1 \times 10^{-5}$. Then, we followed with twice the amount of noise with $\sigma = 2 \times 10^{-5}$ to test both algorithms with an elevated amount of noise. Finally, the average background noise, $\sigma = 8.2 \times 10^{-6}$, was calculated for all images and was used for the last test. The three levels of noise can be seen in **Figure 2**. Since the data is comprised of small values, a scale factor is needed. This factor is calculated using the maximum value found in the dataset as a reference. For both networks, a scale factor of 500 was used.

TABLE 1 | Test results knee single-coil dataset.

σ	Noisy MSE	SURE MSE	Blindspot MSE	NLM MSE
8.2×10^{-6}	6.5954×10^{-11}	3.6943×10^{-11}	3.9075×10^{-11}	3.9826×10^{-11}
1×10^{-5}	9.8777×10^{-11}	4.7123×10^{-11}	4.8734×10^{-11}	4.9732×10^{-11}
2×10^{-5}	4.2101×10^{-10}	9.0616×10^{-11}	8.7264×10^{-11}	9.0004×10^{-11}
σ	**Noisy PSNR**	**SURE PSNR**	**Blindspot PSNR**	**NLM PSNR**
8.2×10^{-6}	28.266	30.866	30.626	30.555
1×10^{-5}	26.512	29.846	29.692	29.610
2×10^{-5}	20.226	27.196	27.329	27.235
σ	**Noisy SSIM**	**SURE SSIM**	**Blindspot SSIM**	**NLM SSIM**
8.2×10^{-6}	0.7238	0.7795	0.7708	0.7661
1×10^{-5}	0.6487	0.7284	0.7215	0.7119
2×10^{-5}	0.3628	0.5579	0.5605	0.5653

TABLE 2 | Test results for the Brainweb dataset.

σ	Noisy MSE	SURE MSE	Blindspot MSE	NLM MSE
50	2,981.044	1,281.977	1,259.961	1,322.726
100	12,332.774	3,508.540	2,758.001	4,059.259
200	50,639.730	9,150.021	7,245.904	11,578.606
σ	**Noisy PSNR**	**SURE PSNR**	**Blindspot PSNR**	**NLM PSNR**
50	33.524	38.015	38.012	37.781
100	27.361	34.036	35.240	33.166
200	21.227	30.301	31.429	28.753
σ	**Noisy SSIM**	**SURE SSIM**	**Blindspot SSIM**	**NLM SSIM**
50	0.7663	0.8971	0.8977	0.8790
100	0.6314	0.8466	0.9066	0.8014
200	0.4710	0.7829	0.8409	0.6996

For the Brainweb dataset, we added three different levels of noise. To understand how the networks behave with different levels of noise, we used low level noise with $\sigma = 50$, middle noise with $\sigma = 100$ and high level noise with $\sigma = 200$. In this case, since the data has much bigger values, a higher sigma is used. The three levels of noise can be seen in **Figure 3**. Note how the data has to be scaled too, specially for the SURE network, which is highly sensitive to the input scale. For the Brainweb dataset, we scaled all input by a factor of 1/25,000. While the blindspot network presented good results even without the scaling factor, it performed slightly better with scaling.

In order to evaluate the proposed algorithm, three quantitative measures were used for the first three tests. Through all tests, a qualitative measure will be used, based on our perception of the images.

The three quantitative measures used are peak signal-to-noise-ratio, PSNR, mean-squared error, MSE web (2020a) and Structural Similarity Index Measure, SSIM Wang et al. (2004). Both MSE and PSNR are used to compare image compression quality, while SSIM is used for measuring the similarity between two images.

MSE represents the cumulative squared error between the compressed and the original image. The lower the value of MSE, the lower the error. MSE can be defined as

$$MSE = \frac{\sum_{M,N} [I_1(m,n) - I_2(m,n)]^2}{M * N} \tag{5}$$

where M and N are the number of rows and columns in the input image.

PSNR computes the peak signal-to-noise ratio between two images. This ratio is used as a quality measurement between

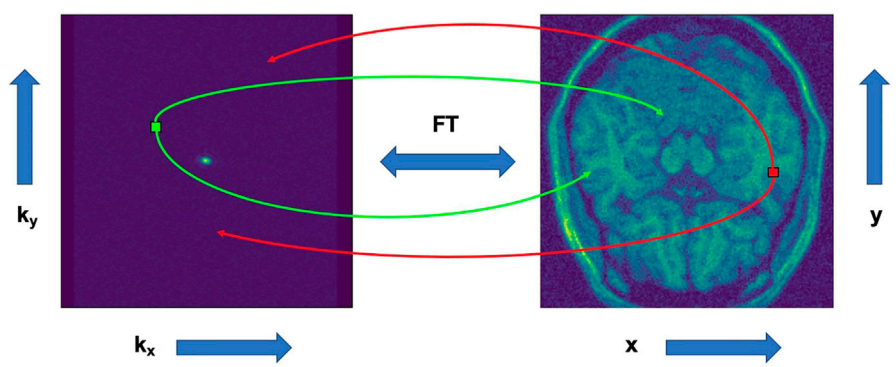

FIGURE 1 | Representation of how points translate between k-space and complex image space.

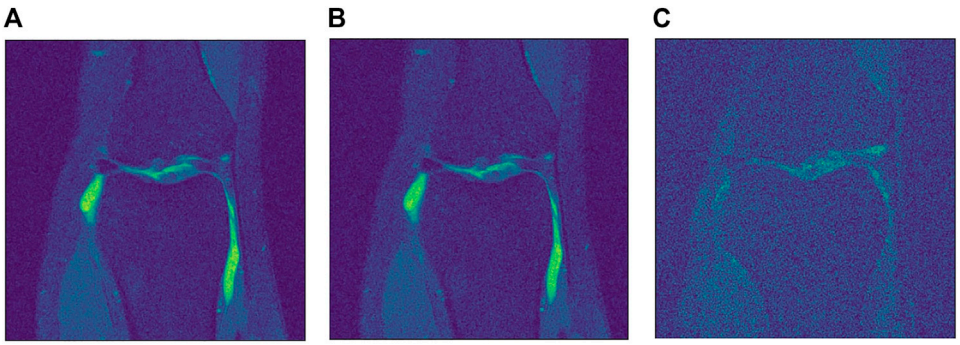

FIGURE 2 | Different levels of noise. **(A)** Low level $\sigma = 8.2 \times 10^{-6}$. **(B)** Medium level $\sigma = 1 \times 10^{-5}$. **(C)** High level $\sigma = 2 \times 10^{-5}$.

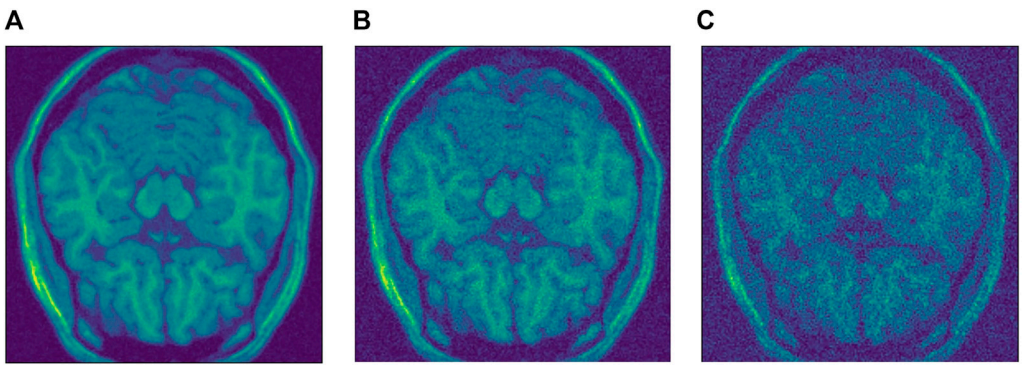

FIGURE 3 | Different levels of noise. **(A)** Low level $\sigma = 50$. **(B)** Medium level $\sigma = 100$. **(C)** High level $\sigma = 200$.

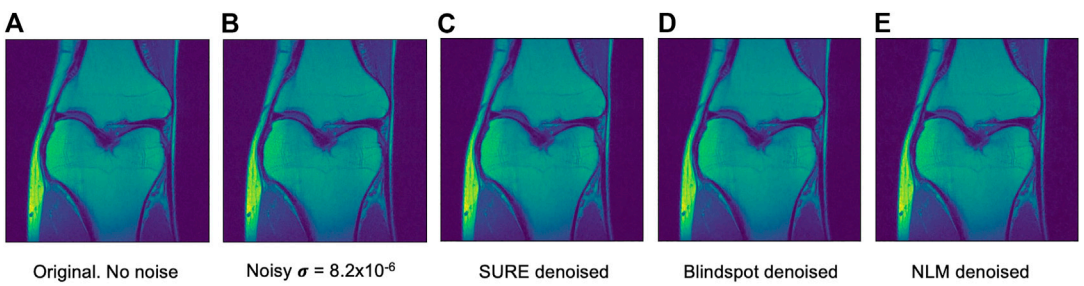

FIGURE 4 | Example of denoised knee MRI for $\sigma = 8.2 \times 10^{-6}$. The example image is the middle slice from one of the subjects. In this case, this is the PSNR for every method for this particular subject. **(A)** Original image, no noise added—**(B)** Noisy image—**(C)** SURE PSNR = 37.092—**(D)** Blindspot PSNR = 37.317—**(E)** NLM PSNR = 36.350.

the original and a compressed or reconstructed image. The higher the PSNR, the better the quality of the image. PSNR can be defined as

$$PSNR = 10\log_{10}\left(\frac{MAX^2}{MSE}\right) \tag{6}$$

where MAX is the maximum achievable value in the input image data type.

SSIM is a method for measuring the similarity between two images. The SSIM index can be viewed as a quality measure of one of the images being compared, taking into account that the other image is regarded as of the ground truth.

The main difference between SSIM and PSNR or MSE is that SSIM quantifies the change in structural information, while PSNR or MSE approach estimate absolute errors. Structural information, such as luminance and contrast, is based on the

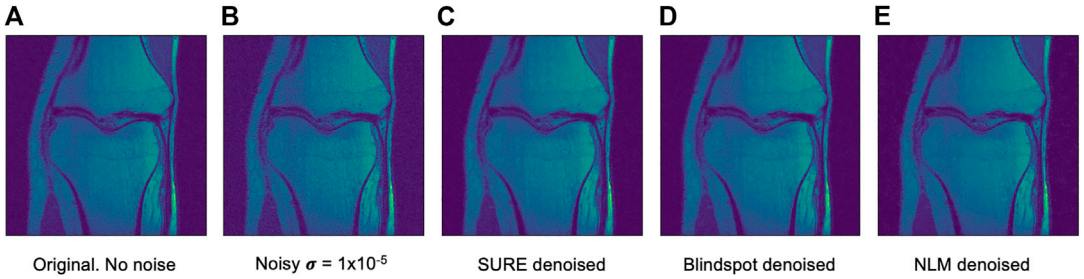

FIGURE 5 | Example of denoised knee MRI for $\sigma = 1 \times 10^{-5}$. The example image is the middle slice from one of the subjects. In this case, this is the PSNR for every method for this particular subject. **(A)** Original image, no noise added—**(B)** Noisy image—**(C)** SURE PSNR = 30.800—**(D)** Blindspot PSNR = 30.953—**(E)** NLM PSNR = 30.189.

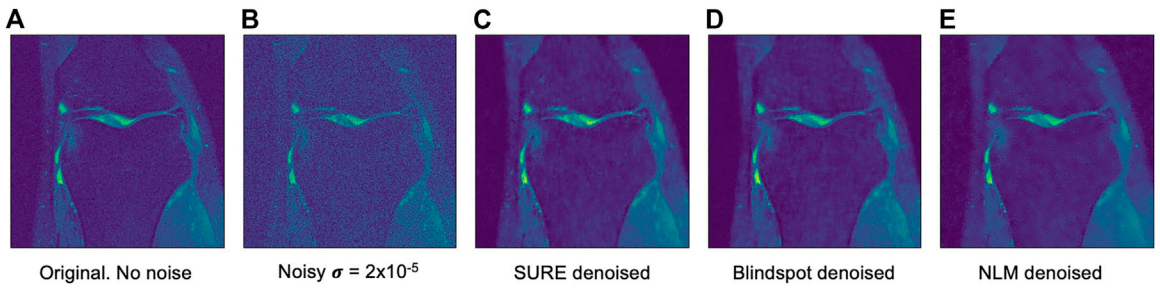

FIGURE 6 | Example of denoised knee MRI for $\sigma = 2 \times 10^{-5}$. The example image is the middle slice from one of the subjects. In this case, this is the PSNR for every method for this particular subject. **(A)** Original image, no noise added—**(B)** Noisy image—**(C)** SURE PSNR = 23.823—**(D)** Blindspot PSNR = 23.931—**(E)** NLM PSNR = 24.086.

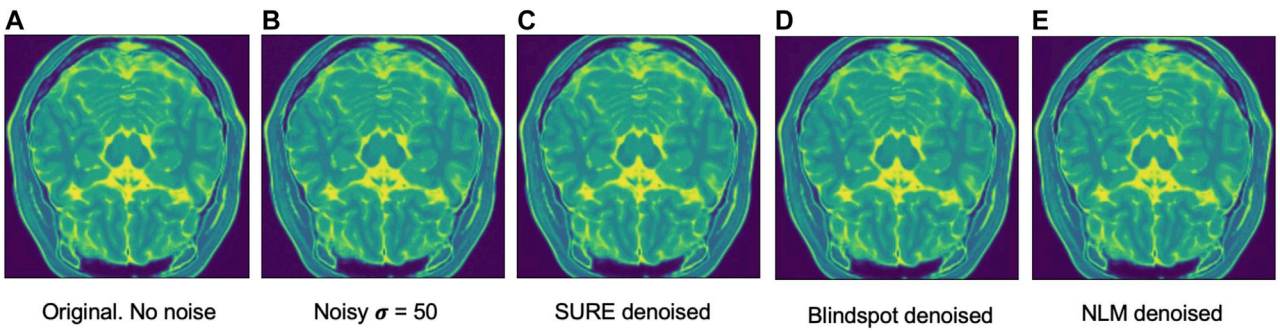

FIGURE 7 | Example of denoised brain MRI for $\sigma = 50$. The example image is the middle slice from one of the subjects. In this case, this is the PSNR for every method for this particular subject. **(A)** Original image, no noise added—**(B)** Noisy image—**(C)** SURE PSNR = 43.883—**(D)** Blindspot PSNR = 44.731—**(E)** NLM PSNR = 43.000.

fact that pixels have inter-dependencies, especially when they are spatially close.

The overall index is a multiplicative combination of the three terms and can be described the following way:

$$\mathrm{SSIM}(x,y) = [l(x,y)]^{\alpha} \cdot [c(x,y)]^{\beta} \cdot [s(x,y)]^{\gamma} \tag{7}$$

where

$$\begin{aligned} l(x,y) &= \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \\ c(x,y) &= \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \\ s(x,y) &= \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \end{aligned} \tag{8}$$

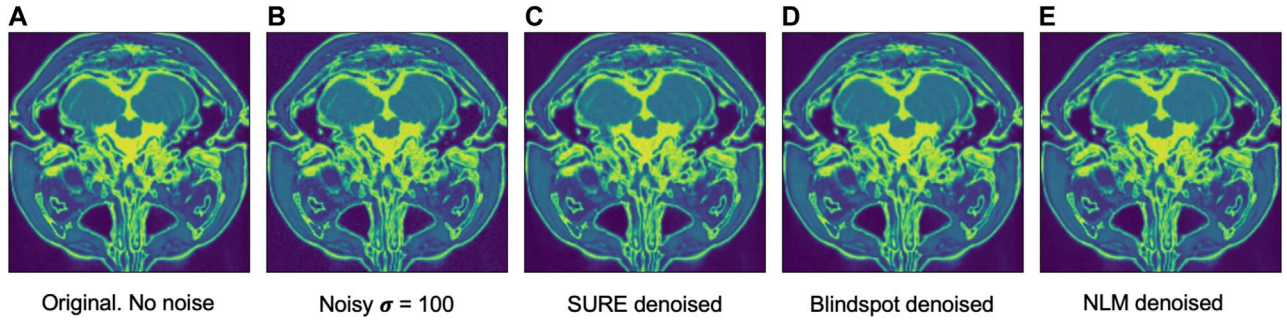

FIGURE 8 | Example of denoised brain MRI for σ = 100. The example image is the middle slice from one of the subjects. In this case, this is the PSNR for every method for this particular subject. **(A)** Original image, no noise added—**(B)** Noisy image—**(C)** SURE PSNR = 38.130—**(D)** Blindspot PSNR = 39.072—**(E)** NLM PSNR = 37.108.

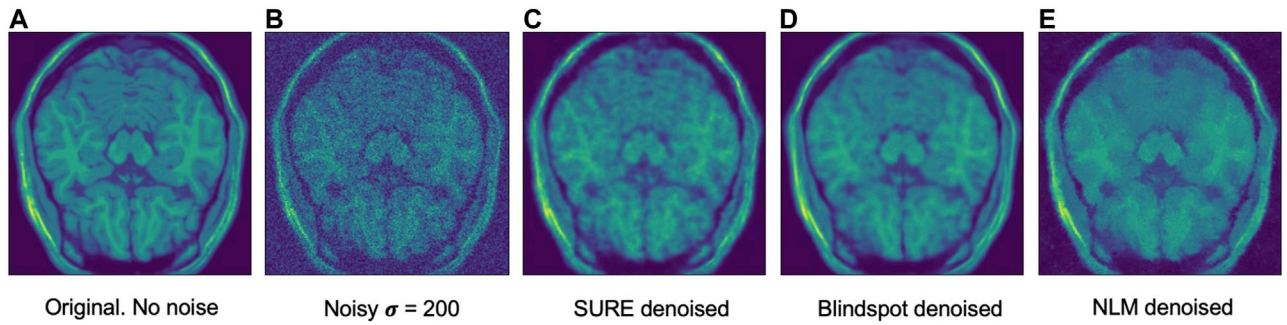

FIGURE 9 | Example of denoised brain MRI for σ = 200. The example image is the middle slice from one of the subjects. In this case, this is the PSNR for every method for this particular subject. **(A)** Original image, no noise added—**(B)** Noisy image—**(C)** SURE PSNR = 29.610—**(D)** Blindspot PSNR = 30.904—**(E)** NLM PSNR = 26.616.

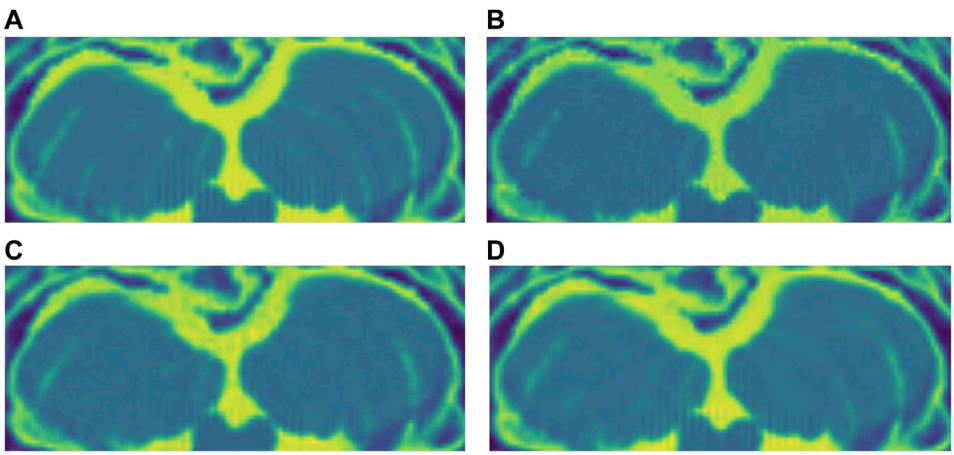

FIGURE 10 | (A) Original close-up. No noise added. **(B)** NLM denoised close-up. **(C)** SURE network denoised close-up. **(D)** Blindspot denoised close-up. Observe how all three algorithms do a good job at denoising, but NLM introduces undesired artifacts.

where μ_x, μ_y, σ_x, σ_y and σ_{xy} are the local means, standard deviations, and cross-covariance for images x, y. If $\alpha = \beta = \gamma = 1$, and $C_3 = C_2/2$ the index simplifies to:

$$\mathrm{SSIM}(x,y) = \frac{\left(2\mu_x\mu_y + C_1\right)\left(2\sigma_{xy} + C_2\right)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)} \tag{9}$$

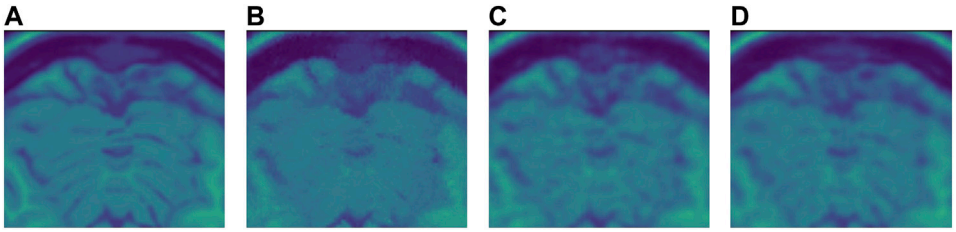

FIGURE 11 | (A) Original close-up. No noise added. **(B)** NLM denoised close-up. **(C)** SURE network denoised close-up. **(D)** Blindspot denoised close-up. Observe how NLM completely removes some of the tissue while both SURE and blindspot, do not remove as much noise, but do a better job at maintaining the tissue's structure without inserting any artifacts.

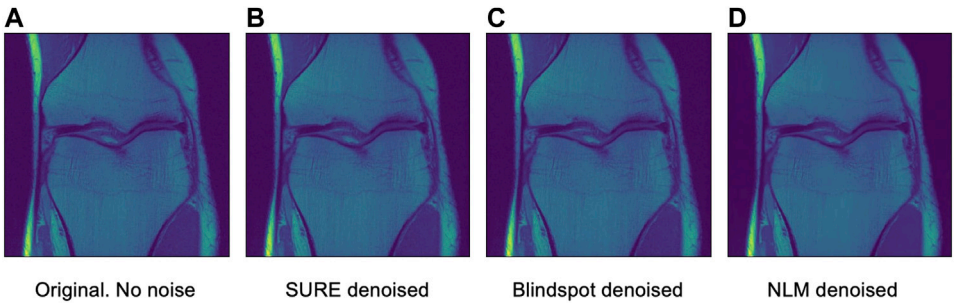

FIGURE 12 | Example 1 of denoised brain MRI without adding any noise. The example image is the middle slice from one of the subjects. **(A)** Original image, no noise—**(B)** SURE denoised image—**(C)** Blindspot denoised image—**(D)** NLM denoised image.

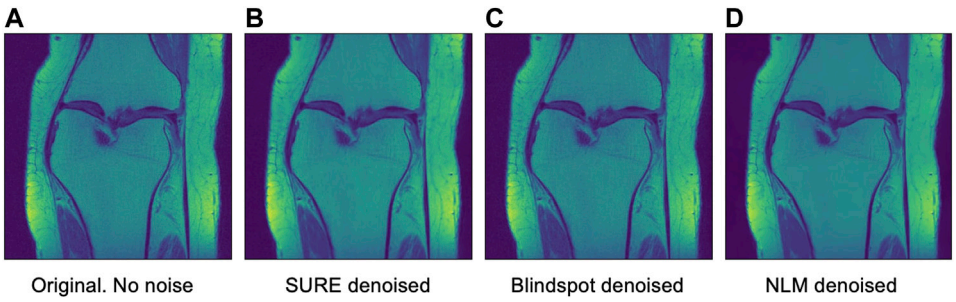

FIGURE 13 | Example 2 of denoised brain MRI without adding any noise. The example image is the middle slice from one of the subjects. **(A)** Original image, no noise—**(B)** SURE denoised image—**(C)** Blindspot denoised image—**(D)** NLM denoised image.

For all the results that are presented here, an optimal h parameter for the NLM algorithm was previously found and set to $h = 0.71$. The patch size was set to 5×5 with a patch distance of 6.

The same tests were done for both the SURE network and the blindspot network, **Table 1, 2** respectively. For each evaluation metric, the best scoring algorithm is highlighted in bold.

4 DISCUSSION

As seen in **Table 1** for the knee data, the SURE network presents better results than NLM and blindspot for both $\sigma = 1 \times 10^{-5}$ and $\sigma = 8.2 \times 10^{-6}$. In both those cases, MSE is smaller and both PSNR and SSIM are larger than NLM and blindspot. Note how in the case of $\sigma = 2 \times 10^{-5}$, NLM does better than the SURE network, but worse than blindspot, except for SSIM. Given that this is an extreme case, where the amount of noise is unrealistically elevated, it would be uncommon to find data in those circumstances.

We can also see that the blindspot network presents better results than NLM for all levels of noise, except for SSIM for $\sigma = 2 \times 10^{-5}$. Compared to SURE, it presents worse results for $\sigma = 1 \times 10^{-5}$ and $\sigma = 8.2 \times 10^{-6}$. Note however, how in the case of $\sigma = 2 \times 10^{-5}$, blindspot outperforms both SURE and NLM except for NLM SSIM. This presents a divergence in the

results previously seen in the complex image space, where for the case of high level noise, NLM was overall better than blindspot and SURE.

For the Brainweb dataset, both networks present better results in all scoring metrics than NLM. The best overall performing network is the blindspot network, edging out the SURE network, except in one case, PSNR for $\sigma = 50$, where SURE is slightly better than blindspot. Again, we believe that in this case both networks do better than NLM even in the presence of high amounts of noise because there is no background noise at all in the original images. Therefore, the networks only need to remove just the added noise.

Another comparison can be done using qualitative measures, based on observing the images and comparing all outputs. Using **Figures 4–6** as references, at a first glance, NLM does a better job at taking noise out, but does it while having a negative effect on the edges and the tissue pixels. NLM does an excellent job when removing noise from the background, but does not do as well on the tissue pixels. This can be a problem, since we want to maintain the tissue structure as much as possible. The SURE network does a better job at preserving the tissue while doing a good job when denoising. In some cases, NLM introduces artifacts that interfere with the tissue pixels. In terms of edge preservation, again NLM presents an undesired effect, which makes the edges look worse than the original image.

For the Brainweb dataset, both networks present better results in all scoring metrics than NLM. The best overall performing network is the blindspot network, edging out the SURE network. We believe that in this case both networks do better than NLM even in the presence of high amounts of noise because there is no background noise at all in the original images. Therefore, the networks only need to remove just the added noise. We can see this in **Figures 7**, **8**, **9**. NLM still presents an undesired effect on the images which can be costly. If we take a closer look, we can see some of the tissue details that the NLM is removing completely and some of the artifacts that it presents. We can clearly see this in **Figures 10**, **11**.

After seeing how both networks outperform NLM in most categories, the next step was to work with the original images from the knee dataset, without adding any extra noise. When doing this test, no quantitative measure can be used, since there is no image to compare to. Therefore, only qualitative measures will be used.

As seen in **Figures 12**, **13**, both networks have mixed results. Both networks still do a better job at preserving the edges and tissue, but sometimes struggle to remove noise from parts of the image without any tissue. This is happening due to a few circumstances. First of all, when training the data, there is no ground truth to compare it to. This can lead to over-training and over-fitting. Second, the inherent noise that the images have, might not be Gaussian noise. This is also supported by the previous results that were obtained for both datasets. Both the SURE and blindspot network were outperformed only in the presence of high levels of noise for the knee dataset. In the same conditions of high level of noise for the Brainweb dataset, both networks outperformed NLM. Therefore, the background noise from the knee dataset has a negative effect on the networks, which might indicate that it is not truly Gaussian. The discrepancy in the type of noise might also be causing the calculated σ to be irrelevant and misleading, since σ is used for both networks. Despite all of this, the networks are competitive with NLM in most cases.

5 CONCLUSION

We evaluated two unsupervised approaches to denoise Magnetic Resonance Image, MRI, one approach based on a Stein's Unbiased Risk Estimator and another one based on a Blindspot network. Using the complex image space, innate to MRI, we tested a real dataset containing knee MRI, and a synthetic dataset consisting of brain MRI. Both networks were compared against Non-Local Means using quantitative and qualitative measures. Both networks outperformed NLM for all scoring metrics except when in the presence of exceptionally high levels of noise. One interesting direction that we would like to explore is 3D denoising using both networks. This is especially compelling for the blindspot network, since we will have to explore a 3D receptive field.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: Need to ask for personalized download code from dataset owners. Requests to access these datasets should be directed to fastmri@med.nyu.edu.

ETHICS STATEMENT

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

MML implemented and ran the experiments and contributed to the manuscript. JF helped organize the datasets and contributed to the manuscript. JV conceived of the project, managed the work and contributed to the manuscript.

FUNDING

This work was supported by the Balsells Foundation and the National Institutes of Health Grant No. 1R15GM128166-01.

REFERENCES

Benou, A., Veksler, R., Friedman, A., and Riklin Raviv, T. (2017). Ensemble of Expert Deep Neural Networks for Spatio-Temporal Denoising of Contrast-Enhanced MRI Sequences. *Med. Image Anal.* 42, 145–159. doi:10.1016/j.media.2017.07.006Accessed November 19, 2020)

Bermudez, C., Plassard, A. J., Davis, T. L., Newton, A. T., Resnick, S. M., and Landman, B. A. (2018). Learning Implicit Brain MRI Manifolds with Deep Learning. *Proc. SPIE Int. Soc. Opt. Eng.* 10574, 408–414. doi:10.1117/12.2293515Accessed November 19, 2020)

Bottou, L. (1999). *On-Line Learning and Stochastic Approximations*. USA: Cambridge University Press, 9–42. doi:10.1017/cbo9780511569920.003

Buades, A., Coll, B., and Morel, J. (2005). "A Non-local Algorithm for Image Denoising," in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 60–65.

Cocosco, C. A., Kollokian, V., Kwan, R. K.-S., and Evans, A. C. (1997). Brainweb: Online Interface to a 3D MRI Simulated Brain Database. *NeuroImage* 5, 4. doi:10.1016/S1053-8119(97)80018-3

Collins, D. L., Zijdenbos, A. P., Kollokian, V., Sled, J. G., Kabani, N. J., Holmes, C. J., et al. (1998). Design and Construction of a Realistic Digital Brain Phantom. *IEEE Trans. Med. Imaging* 17, 463–468. doi:10.1109/42.712135

Dabov, K., Foi, A., Katkovnik, V., and Egiazarian, K. (2007). Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering. *IEEE Trans. Image Process.* 16, 2080–2095. doi:10.1109/TIP.2007.901238Accessed November 19, 2020)

Eun, D.-i., Jang, R., Ha, W. S., Lee, H., Jung, S. C., and Kim, N. (2020). Deep-learning-based Image Quality Enhancement of Compressed Sensing Magnetic Resonance Imaging of Vessel wall: Comparison of Self-Supervised and Unsupervised Approaches. *Sci. Rep.* 10, 13950. doi:10.1038/s41598-020-69932-wAccessed November 28, 2020)

Gal, Y., Mehnert, A. J. H., Bradley, A. P., McMahon, K., Kennedy, D., and Crozier, S. (2010). Denoising of Dynamic Contrast-Enhanced MR Images Using Dynamic Nonlocal Means. *IEEE Trans. Med. Imaging* 29, 302–310. doi:10.1109/TMI.2009.2026575Accessed November 19, 2020)

Jiang, D., Dou, W., Vosters, L., Xu, X., Sun, Y., and Tan, T. (2018). Denoising of 3D Magnetic Resonance Images with Multi-Channel Residual Learning of Convolutional Neural Network. *Jpn. J. Radiol.* 36, 566–574. doi:10.1007/s11604-018-0758-8Accessed November 28, 2020)

Kingma, D. P., and Ba, J. (2015). "Adam: a Method for Stochastic Optimization," in 3rd International Conference on Learning Representations, ICLR 2015 (San Diego, CA, USA: Conference Track Proceedings).

Knoll, F., Zbontar, J., Sriram, A., Muckley, M. J., Bruno, M., Defazio, A., et al. (2020). fastMRI: a Publicly Available Raw K-Space and Dicom Dataset of Knee Images for Accelerated MR Image Reconstruction Using Machine Learning. *Radiol. Artif. Intell.* 2, e190007. doi:10.1148/ryai.2020190007Accessed November 27, 2020)

Krull, A., Buchholz, T. O., and Jug, F. (2018). *Noise2void - Learning Denoising from Single Noisy Images*. CoRR abs/1811, 10980.

Kwan, R. K.-S., Evans, A. C., and Pike, G. B. (1996). An Extensible MRI Simulator for post-processing Evaluation. *Visualization Biomed. Comput.* VBC'96, 135–140. doi:10.1007/BFb0046947

Kwan, R. K.-S., Evans, A. C., and Pike, G. B. (1999). Mri Simulation-Based Evaluation of Image-Processing and Classification Methods. *IEEE Trans. Med. Imaging* 18, 1085–1097. doi:10.1109/42.816072

Laine, S., Karras, T., Lehtinen, J., and Aila, T. (2019). "High-quality Self-Supervised Deep Image Denoising," in *Advances in Neural Information Processing Systems* (Curran Associates, Inc.), 6970–6980. (Accessed November 19, 2020)

Lehtinen, J., Munkberg, J., Hasselgren, J., Laine, S., Karras, T., Aittala, M., et al. (2018). "Noise2Noise: Learning Image Restoration without Clean Data," in Proceedings of the 35th International Conference on Machine Learning, 2965–2974. (Accessed November 27, 2020)

Mohan, J., Krishnaveni, V., and Guo, Y. (2014). A Survey on the Magnetic Resonance Image Denoising Methods. *Biomed. Signal Process. Control.* 9, 56–69. doi:10.1016/j.bspc.2013.10.007Accessed November 19, 2020)

Ramani, S., Blu, T., and Unser, M. (2008). Monte-Carlo SURE: a Black-Box Optimization of Regularization Parameters for General Denoising Algorithms. *IEEE Trans. Image Process.* 17, 1540–1554. doi:10.1109/TIP.2008.2001404Accessed November 19, 2020)

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention* (MICCAI), 234–241. doi:10.1007/978-3-319-24574-4_28Accessed November 27, 2020)

Sardy, S., Tseng, P., and Bruce, A. (2001). Robust Wavelet Denoising. *IEEE Trans. Signal. Process.* 49, 1146–1152. doi:10.1109/78.923297Accessed November 28, 2020)

Soltanayev, S., and Chun, S. Y. (2018). "Training Deep Learning Based Denoisers without Ground Truth Data," in *Advances in Neural Information Processing Systems* (Curran Associates, Inc.), 3257–3267. doi:10.5555/3327144.3327246Accessed November 19, 2020)

Stein, C. M. (1981). Estimation of the Mean of a Multivariate normal Distribution. *Ann. Statist.* 9, 1135–1151. doi:10.1214/aos/1176345632Accessed November 19, 2020)

Tomasi, C., and Manduchi, R. (1998). "Bilateral Filtering for gray and Color Images," in Proceedings of the Sixth International Conference on Computer Vision (IEEE Computer Society), ICCV '98), 839.

Tripathi, P. C., and Bag, S. (2020). Cnn-dmri: a Convolutional Neural Network for Denoising of Magnetic Resonance Images. *Pattern Recognition Lett.* 135, 57–63. doi:10.1016/j.patrec.2020.03.036Accessed November 28, 2020)

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., and Manzagol, P. A. (2010). Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion. *J. Mach. Learn. Res.* 11, 3371–3408. doi:10.5555/1756006.1953039Accessed November 19, 2020)

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image Quality Assessment: from Error Visibility to Structural Similarity. *IEEE Trans. Image Process.* 13, 600–612. doi:10.1109/TIP.2003.819861 Accessed November 19, 2020)

web (1998). *Online* (Accessed November 19, 2020)

web (2017). *Online* (Accessed November 28, 2020)

web (2020a). *Online* (Accessed November 19, 2020)

web (2020b). *Online* (Accessed November 28, 2020)

Xu, J., Huang, Y., Cheng, M.-M., Liu, L., Zhu, F., Xu, Z., et al. (2020). Noisy-as-clean: Learning Self-Supervised Denoising from Corrupted Image. *IEEE Trans. Image Process.* 29, 9316–9329. doi:10.1109/TIP.2020.3026622Accessed November 28, 2020)

Zbontar, J., Knoll, F., Sriram, A., Muckley, M. J., Bruno, M., Defazio, A., et al. (2018). *Fast MRI: An Open Dataset and Benchmarks for Accelerated MRI*.

Zhang, K., Zuo, W., Chen, Y., Meng, D., and Zhang, L. (2017). Beyond a Gaussian Denoiser: Residual Learning of Deep Cnn for Image Denoising. *IEEE Trans. Image Process.* 26, 3142–3155. doi:10.1109/TIP.2017.2662206Accessed November 28, 2020)

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Big Data

ORIGINAL RESEARCH
published: 16 July 2021
doi: 10.3389/fdata.2021.568352

An Integrated Deep Network for Cancer Survival Prediction Using Omics Data

Hamid Reza Hassanzadeh[1] *and May D. Wang*[2]*

[1]School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA, United States, [2]Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, GA, United States

As a highly sophisticated disease that humanity faces, cancer is known to be associated with dysregulation of cellular mechanisms in different levels, which demands novel paradigms to capture informative features from different omics modalities in an integrated way. Successful stratification of patients with respect to their molecular profiles is a key step in precision medicine and in tailoring personalized treatment for critically ill patients. In this article, we use an integrated deep belief network to differentiate high-risk cancer patients from the low-risk ones in terms of the overall survival. Our study analyzes RNA, miRNA, and methylation molecular data modalities from both labeled and unlabeled samples to predict cancer survival and subsequently to provide risk stratification. To assess the robustness of our novel integrative analytics, we utilize datasets of three cancer types with 836 patients and show that our approach outperforms the most successful supervised and semi-supervised classification techniques applied to the same cancer prediction problems. In addition, despite the preconception that deep learning techniques require large size datasets for proper training, we have illustrated that our model can achieve better results for moderately sized cancer datasets.

Keywords: deep belief networks, integrated cancer survival analysis, RNA-seq, precision medicine, deep learning, multi-omics

OPEN ACCESS

Edited by:
Shuai Li,
Swansea University, United Kingdom

Reviewed by:
Indu Khatri,
Leiden University Medical Center, Netherlands
Ricardo De Matos Simoes,
Dana–Farber Cancer Institute, United States
Ameer Tamoor Khan,
Hong Kong Polytechnic University, Hong Kong

***Correspondence:**
May D. Wang
maywang@gatech.edu

Specialty section:
This article was submitted to Medicine and Public Health, a section of the journal Frontiers in Big Data

Received: *01 June 2020*
Accepted: *01 June 2021*
Published: *16 July 2021*

Citation:
Hassanzadeh HR and Wang MD (2021) An Integrated Deep Network for Cancer Survival Prediction Using Omics Data. Front. Big Data 4:568352. doi: 10.3389/fdata.2021.568352

INTRODUCTION

Advances in big data and high-throughput technologies during the past decade have led to massive accumulation of high-dimensional omics data, which enables the data-driven prediction of disease prognosis using molecular profiles. However, this data-driven prognosis remains challenging because of the interplay of mostly unknown molecular factors from a haystack of millions of molecular features. The general practice in prognosis of most of the malign diseases has been based on the traditional methods without a comprehensive analysis of genetic and molecular profiles. This is primarily due to the lack of reliable clinical decision support systems (CDSSs) that can efficiently model and integrate information into actionable knowledge.

The association of molecular profiles with the onset of chronic diseases and their sub-types and prognoses has been extensively reviewed and reported during the past years (Hsieh et al., 2018; Collisson et al., 2019; Sicklick et al., 2019). Despite the success of a number of these approaches, majority of them utilize the so-called shallow-learners, which often fall short in learning higher-order abstract representations of the data and fail to capture complex inter-modality or intra-modality interactions of features or their relationship with respect to clinical endpoints of interest. Often,

shallow learners use a limited set of features derived from the expert knowledge or feature reduction techniques, such as the principal component analysis (PCA). Thus, they are limited in their ability to learn non-linear higher-level informative features. In contrast, deep learning (LeCun et al., 2015) is revolutionizing the field of feature learning (also known as representation learning) in biomedicine (Alipanahi et al., 2015; Fan et al., 2015; Park and Kellis, 2015; Spencer et al., 2015; Wang S. et al., 2016). Inspired by neuroscience, the power of deep learning is its ability to represent high-dimensional data by multiple levels of non-linearity abstractions and to train DL models with more effective optimizations and regularization techniques. Once such a representation is derived, any classifier for the prediction task can use it.

To date, some studies have designed deep methods for prediction and prognosis of cancer using different types of modalities. Fakoor et al. (2020) used a stack of sparse auto-encoders along with an augmenting dimensionality reduction step using PCA, to learn features from gene expression data that can help classify cancer types. They developed three variants of their proposed paradigm and showed that they perform reasonably well across different datasets in some of their devised experiments, but not all. The addition of PCA to extract new features from randomly selected probes is a necessary step in their pipeline as the sparse stacked auto-encoder is not enough by itself to predict informative features. Moreover, their approach uses only a single data modality, i.e., gene expression data, for prediction of cancer type. In another study, Kumar et al. (2015) used a similar approach to Fakoor et al. (2020), in their own domain of interest, to create useful features from CT images to classify benign vs. malignant lung nodules. They showed that their approach resulted in a performance boost compared to the state-of-the-art approaches. Azizi (2020) developed a supervised pipeline, based on the deep belief network (DBN), for detection of prostate cancer given ultrasound temporal data. The author used deep belief networks to learn useful features, which are then fed into a support vector machine classifier to predict cancer. In another study, Liang et al. (2015) integrated several restricted Boltzmann machines (RBMs) for an unsupervised task of grouping cancer tumors into different clusters using cross-platform but same-type molecular data. They showed that patients grouped in different clusters exhibit differentiable Kaplan–Meier survival curves, which is an indication of the soundness of their proposed clustering approach. More recently, Zeng et al. (2020) used a supervised learning approach based on the convolutional neural network for subtyping of breast cancer. Besides the supervised nature of the proposed model there, CNNs are severely restricted in capturing long distance relations, due to their short receptive fields, especially when the number of input features is orders of magnitude larger than the utilized kernel width.

Despite many DL applications in different biomedical areas, their success in cancer prediction and prognosis is still limited. This is because deep architectures require high volumes of labeled data samples (due to their expressiveness, Hastie et al., 2009) to train DL models without data overfitting, which is a requirement not always met in cancer-related domains. In this study, we develop an integrated semi-supervised deep learning for risk prediction in cancer cohorts with patients' molecular profiles. We present an integrated deep architecture to predict cancer survival given the molecular profiles of cancer tumors. We show that our integrated deep model can leverage the available unlabeled data to enhance learning our deep model, a task that is often achieved using semi-supervised learning frameworks. Furthermore, we illustrate that the proposed pipeline outperforms the support vector machine (SVM), a supervised learner that has been successfully used in cancer-related domains (Kim et al., 2012a; Ahmad, 2013; Tseng et al., 2014) as well as the Laplacian SVM, an important graph-based semi-supervised learning paradigm that is promising in solving similar problems (Kim et al., 2012b; Kim and Shin, 2013; Park et al., 2013).

DATASETS

The Cancer Genome Atlas (TCGA) (Weinstein et al., 2013) data portal in the NCI/NIH (Cerami et al., 2012; Wang Z. et al., 2016) hosts multi-modality data of thousands of patients. In this study, we used data from kidney renal clear cell carcinoma (KIRC) and head and neck squamous cell carcinoma (HNSC) diseases from the TCGA data bank. We selected KIRC and HNSC because they are moderately sized. On the one hand, they are not too small[1] in the number of specimens profiled, and on the other hand, we did not intend to select a cancer type with a relatively large number of samples, such as the invasive breast carcinoma, to showcase the efficacy of the utilized architecture in learning generalizable models. Furthermore, we downloaded RNA-seq expression profiles of patients suffering from neuroblastoma (NB) pediatric cancer from a previously published study (Zhang et al., 2015). For the RNA-seq expression profiles, we used three data modalities per sample, namely, the gene, the isoform, and the junction. For the KIRC and HNSC datasets, these were produced by Illumina HiSeq 2000 platforms and quantified by RSEM (Li and Dewey, 2011). In case of the NB dataset, we selected the results of mapping the reads to the AceView (Thierry-Mieg and Thierry-Mieg, 2006) annotation through the Magic alignment tool (Thierry-Mieg and Thierry-Mieg, 2006). We also used the miRNA expression profiles for the KIRC and HNSC datasets, which were generated by the Illumina GAIIx platform, and finally, the Illumina Infinium HumanMethylation27 platform produced the DNA methylation data for the KIRC disease only. **Table 1** lists the available modalities and their statistics for each dataset.

MATERIALS AND METHODS

Recent years have witnessed a surge of interest in deep learning (DL) and its successful applications in different domains (see

[1]Even though DBNs partially alleviate data scarcity in training, when the dataset is too small, they cause data overfitting.

TABLE 1 | Data description.

Data modality (platform)		Dataset	# of features	# of available samples		
				Labeled		Unlabeled
				Pos.	Neg.	
RNA-seq (Illumina HiSeq 2000)	Gene	KIRC	20533	110	141	281
		HNSC	20533	115	128	276
		NB	60780	115	104	279
	Isoform	KIRC	73601	110	141	281
		HNSC	73601	115	128	276
		NB	263546	115	104	279
	Junction	KIRC	249579	110	141	281
		HNSC	249579	115	128	276
		NB	340416	115	104	279
miRNA (Illumina GAIIx 2000)		KIRC	1048	106	150	269
		HNSC	1048	116	130	276
Methylation (Illumina Infinium HumanMethylation27)		KIRC	21403	111	142	520

promising examples in Hassanzadeh and Wang 2016; Esteva et al., 2017; Hassanzadeh et al., 2017; Miotto et al., 2017), such as image processing, speech recognition, computer vision, and more recently in biology. Despite its success in a wide range of areas, there are only a handful of studies reporting success stories about the application of DL in cancer-related domains. In fact, several attempts to deploy DL in biomedical domains have failed to outperform other classical methods (Fakoor et al., 2020; Ditzler et al., 2015). This is due to the selection of wrong components and the DL architectures for the selected tasks. Moreover, these pipelines are often designed for supervised tasks, which are inefficient when dealing with censored data that are prevalent in cancer databases. In this study, we developed a deep learning model to deal with the dataset size limitation. This strategy is equivalent to the semi-supervised learning (SSL) strategy, where we leverage unlabeled samples to guide the training process of network weights. Until recently, semi-supervised learning (SSL) (Chapelle et al., 2009) approaches have been the dominant practice to learn models that use both labeled and unlabeled data. This is mainly due to their higher performance compared to the purely supervised or unsupervised techniques. Different SSL paradigms try to take advantage of unlabeled data in different ways, but they all capture the probability distribution of the input samples either directly or indirectly. In other words, what gives SSL techniques an advantage over the supervised methods is their ability to exploit all data, irrespective of the labels, to model a more realistic marginal distribution of the input.

Data Description

Among all SSL techniques, Laplacian support vector machine (LapSVM) is an outstanding recent technique that falls under the category of graph-based SSL paradigms, which builds a graph representation of the data (labeled and unlabeled) based on domain knowledge or the similarity among samples. It has shown the state-of-the-art performance in semi-supervised classification problems (Belkin et al., 2006; Melacci and Belkin, 2011). The underlying assumption in LapSVM is that the marginal distribution of the data can be represented in a low-dimensional manifold that is representable by a similarity graph. Formally speaking, if the marginal distribution of the data can be supported on a low-dimensional Riemannian manifold, then by exploiting its intrinsic geometry through enforcing a smoothness constraint, one can introduce a preferential bias in the learning process to yield a more accurate model. Thus, by adding a new regularizer term for the smoothness on the manifold, one can expand the framework of supervised learning methods that are fully described by a cost function and regularizers such as SVM and ridge regression to exploit the structure of the data using both the labeled and the unlabeled data. Consequently, the Laplacian SVM solution is defined as

$$f^* = \underset{f \in \mathcal{H}_K}{\operatorname{argmin}} \frac{1}{l} \sum_{i=1}^{l} \gamma_A \|f\|_K^2 + \gamma_I + \int_{x \in \mathcal{M}} \|\nabla_{\mathcal{M}} f\|^2 d\mathcal{P}_X(x) \quad (1)$$

where V is the cost function, γ_A, γ_I are the regularizer coefficients in the so-called ambient and the manifold spaces, respectively, $\mathcal{P}_X(x)$ is the marginal distribution of the data, and $\mathcal{H}_K$ is the corresponding reproducing kernel space. Belkin et al. (2006) showed that, under certain conditions, the term corresponding to the manifold regularization can be approximated with $\frac{\gamma_I}{N^2} \boldsymbol{f}^T L \boldsymbol{f}$, where N is the number of samples, $\boldsymbol{f} = [f(x_1), \ldots, f(x_N)]^T$, and L is the Laplacian of the graph underlying the data. As a result, solvers that optimize the supervised SVM problems efficiently can be readily used to find the solution to the semi-supervised LapSVM problem too.

Restricted Boltzmann Machines

RBMs (Hinton and Salakhutdinov, 2006) are the most common building blocks in deep probabilistic models such as DBNs (Goodfellow et al., 2016). These are undirected probabilistic graphical models with a fully bipartite graphical structure (see **Figure 1A**) that contains a layer of visible units, v, and a layer of latent variables, h. Due to the expressiveness of these models, they have become popular techniques in learning features that are represented by the latent layers. RBMs can also be stacked on top of each other to make deeper architectures. Each unit in an RBM is a binary random variable, and the visible layer of the first RBM in the stack

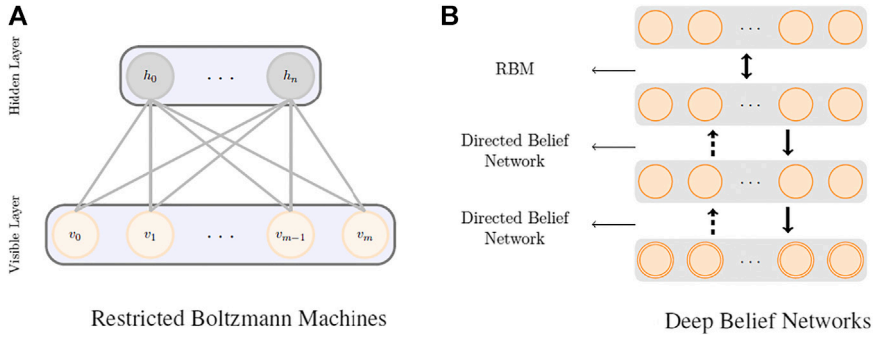

FIGURE 1 | RBM **(A)** and DBN **(B)** model architectures.

represents the input data. The joint probability distribution in an RBM is modeled as

$$P(\boldsymbol{v},\boldsymbol{h}) = \frac{1}{Z}\exp(-E(\boldsymbol{v},\boldsymbol{h})), \tag{2}$$

where $E(\boldsymbol{v},\boldsymbol{h}) = -\boldsymbol{b}^T\boldsymbol{v} - \boldsymbol{c}^T\boldsymbol{h} - \boldsymbol{v}^T\boldsymbol{W}\boldsymbol{h}$ is the energy function, $\boldsymbol{Z}$ is the partition function that normalizes the energy, and W is the weight matrix that characterizes the underlying model. Despite the intractable nature of the joint distribution due to the partition function in **Eq. 2**, $P(\boldsymbol{v},\boldsymbol{h}) = \frac{1}{Z}\exp(-E(\boldsymbol{v},\boldsymbol{h}))$, the conditional distributions are factorial in nature, that is,

$$P\left(h_j = 1 \middle| \boldsymbol{v}\right) = \sigma\left(c_j + \boldsymbol{v}^T\boldsymbol{W}_{:j}\right),$$
$$P\left(v_j = 1 \middle| \boldsymbol{h}\right) = \sigma\left(b_j + \boldsymbol{h}^T\boldsymbol{W}_{j:}\right),$$

where σ is the logistic sigmoid function. This makes the overall distribution amenable to sampling, and hence, efficient estimation of the joint probability distribution under the model can be made.

Deep Belief Networks

The DBN was one of the first attempts that outperformed the state-of-the-art shallow learners in image processing and marks the beginning of the deep learning revolution. Even though this class of deep models do not perform as well as the more advanced deep models when a large body of labeled data is available, they do surprisingly well in circumstances with less data.

Figure 1B shows a schematic representation of a deep belief network. DBNs are generative models formed by stacking several directed belief networks trying to capture causal relations and an RBM layer on the top that acts as an associative model. The joint probability distribution for a DBN with l layers is given by

$$P\left(\boldsymbol{x} = \boldsymbol{h}^{(0)}, \boldsymbol{h}^{(1)}, \ldots, \boldsymbol{h}^{(l)}\right) = \left(\prod_{k=0}^{l-2} P\left(\boldsymbol{h}^k \middle| \boldsymbol{h}^{k+1}\right)\right) P\left(\boldsymbol{h}^{(l)}, \boldsymbol{h}^{(l-1)}\right),$$
$$P\left(\boldsymbol{h}^{(l)}, \boldsymbol{h}^{(l-1)}\right) \propto \exp\left(b^{(l)}\boldsymbol{h}^{(l)} + b^{(l-1)^T}\boldsymbol{h}^{(l-1)} + \boldsymbol{h}^{(l-1)^T}\boldsymbol{W}\boldsymbol{h}^{(l)}\right),$$
$$P\left(\boldsymbol{h}^k \middle| \boldsymbol{h}^{k+1}\right) = \sigma\left(b_i^{(k)} + \boldsymbol{W}_{:,i}^{(k+1)^T}\boldsymbol{h}^{(k+1)}\right) \forall i, k \in 1, \ldots, l-2,$$

where $\boldsymbol{b}^{(l)}$, $\boldsymbol{W}^{(l)}$ are the bias vector and the weight vector for the lth layer, respectively.

Thus, DBNs provide multi-layer probabilistic representations of data in an unsupervised way, and as a result, latent representation of the low-level features can be obtained using several levels of abstraction. Training and inference in deep belief nets is not a tractable task. We adopt a heuristic approach called the contrastive divergence (CD-k) proposed by Hinton and Salakhutdinov (2006) to do the training and inference in our model. In summary, this approach begins with training an RBM to maximize $\mathbb{E}_{v \sim p_{data}}$ and then another RBM to approximately maximize $\mathbb{E}_{\boldsymbol{v} \sim p_{data}}\mathbb{E}_{\boldsymbol{h}^{(1)} \sim p^{(1)}(\boldsymbol{h}^{(1)}|v)} \log p^{(2)}(\boldsymbol{h}^{(1)})$, where $p^{(1)}$and $p^{(2)}$ are the probability distributions characterized by the first and the second RBMs, respectively. In other words, the second RBM is trained to model the distribution over its input derived from sampling the first RBM. This process can be repeated for as many layers as needed and increases the variational lower bound on the log-likelihood of the data each time a new layer is added. The DBN initializes the weights of multi-layer perceptrons (MLPs), a procedure dubbed as pre-training (Hinton and Salakhutdinov, 2006), to set the stage for the fine-tuning phase in the next step. Specifically, by adding a sigmoid layer on top of a DBN and reusing the generatively trained weights as the initial weights, we can discriminatively train the underlying MLP (Bengio, 2007) via conventional back-propagation–based techniques to converge to a more accurate local optimum. Pre-training differentiates itself from the SSL techniques by finding a proper initial point within the complex search space in an informed way, without modifying the objective function (Erhan, 2010).

Model Architecture

Figure 2 depicts the architecture of the proposed model. First, the patients' overall survival statuses are retrieved from the clinical data in TCGA. Patients in the KIRC, HNSC, and NB datasets who at the time of their last follow-up had survived for at least 5, 2.5, and 9 years, respectively, were assigned to the positive survival class. Similarly, patients who did not survive for the corresponding period of time were assigned to the negative (deceased) class, and the rest, i.e., patients whose latest statuses were known to be alive and who yet did not live with their disease long enough to pass the selected threshold, were put into the unlabeled set. **Table 1** demonstrates the number of positive, negative, and, unlabeled patients. With each of the

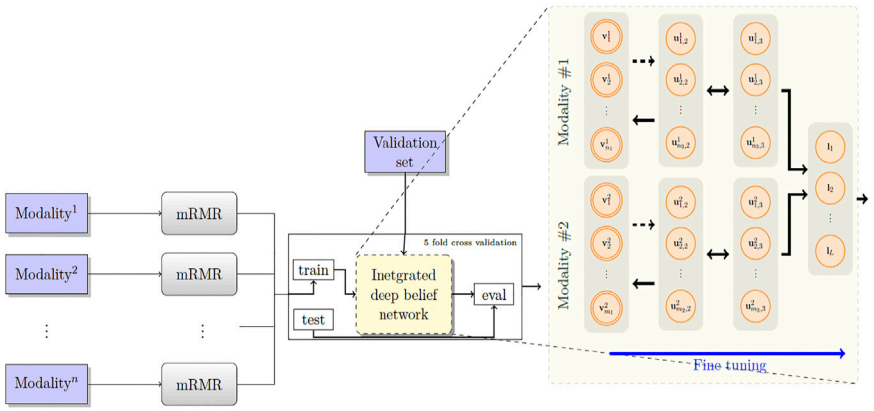

FIGURE 2 | The proposed model RBM. A set of features are first selected for each molecular profile, using mRMR. Then, for each molecular profile, latent features are derived using deep belief nets, which are then fed into a sigmoid layer for downstream prediction.

datasets, 15% of the labeled samples randomly selected to be a validation dataset. The remaining 85% samples in the labeled set are put into five-folds to conduct a five-fold cross-validation for later analysis of our proposed pipeline. Next, we use mRMR (Peng et al., 2005) to reduce the dimensionality of input modalities so that the uninformative features are removed. mRMR is an incremental search algorithm that looks for a subset of features with the highest relevance to the class labels and lowest redundancy compared to each other. To select the most relevant and yet least redundant probes from the underlying molecular profiles, we discretized the scores pertaining to each probe into three bins, based on its standard deviation across samples (i.e., $(-\infty, -0.7\sigma]$, $(-0.7\sigma,\ 0.7\sigma)$, and $[0.7\sigma, \infty)$). We then picked the set of top 50 mRMR selected probes (in a non-discretized form) from all modalities combined, excluding the microRNA, as well as the top 20 probes from the microRNA profiles. Subsequently, we computed the per-probe z-score of the resulting subset of features before feeding them into our models. It is worth noting that as our DBN-based models are theoretically capable of extracting higher-order informative features from a pool of raw input features, selecting an optimal number of features from the molecular profiles is not a concern here, as long as we choose a proportionate number of input features.

Next, we built a network with two hidden layers (the first layer being a belief net comprising 15 neurons and the second layer is a restricted Boltzmann machine with another 15 neurons). We trained the model corresponding to each modality using the contrastive divergence algorithm with k set to one, i.e., CD-1, used the stochastic gradient descent with a batch size of 25 and a weight decay of 0.001, and continued pre-training for 3,000 epochs to train the network. Subsequently, we augmented our probabilistic DBN model with an additional fully connected sigmoid layer followed by a softmax layer and initialized the weights of the previous layers with those found by the CD-1 algorithm, as this has been shown to be a valuable initialization for such networks. Furthermore, we used our labeled data in the training set to fine-tune the model with a maximum of 500 epochs according to an early-stopping training strategy.

Because cancer has been known to be the outcome of dysregulation of cellular mechanisms in different levels, a single molecular data modality may not adequately explain the sophisticated underlying mechanisms. To account for the interactions, or otherwise correlations between molecular factors with respect to the endpoint we are exploring (which is the risk category of patients), we formed a hybrid model by fusing the intermediate-level features (i.e., features that were generated before the softmax layers) for pairs[2] of single-modality models and stacked a softmax layer on top of them (see **Figure 2**). We also explored different model spaces by adding more layers on top of the fusion layer as it theoretically could result in capturing more intricate interactions and hence better performance gains, and we found that such architectures do not bring about further improvements, which can be explained by the limited size of our training sets and the complexity of the task. Finally, we trained the overall model end-to-end, using the cross-entropy loss and the stochastic gradient descent optimizer.

RESULTS

In this study, we investigate two major questions. First, would a deep classifier help improving the performance of single-modality models in predicting survivals? Second, would the integrated deep belief net outperform the single-modality models? Positive answers to these questions would support the applicability of deep networks in predicting survival and the feasibility of DBNs in utilizing the redundant intermediate features to boost the prediction performance. We compared the performance of the proposed model with two baselines: 1) when we substitute the deep belief parts with the supervised support vector machine (SVM) classifiers and 2) when we use semi-supervised graph-based Laplacian SVMs as a surrogate method. To address the

[2]We tried the fusion of more than two data modalities as well; however, this led to model overfitting and a drop in accuracy, due to the small size of datasets. We hypothesize that adding more modalities can lead to a synergistic boost in the prediction performance if more data become available.

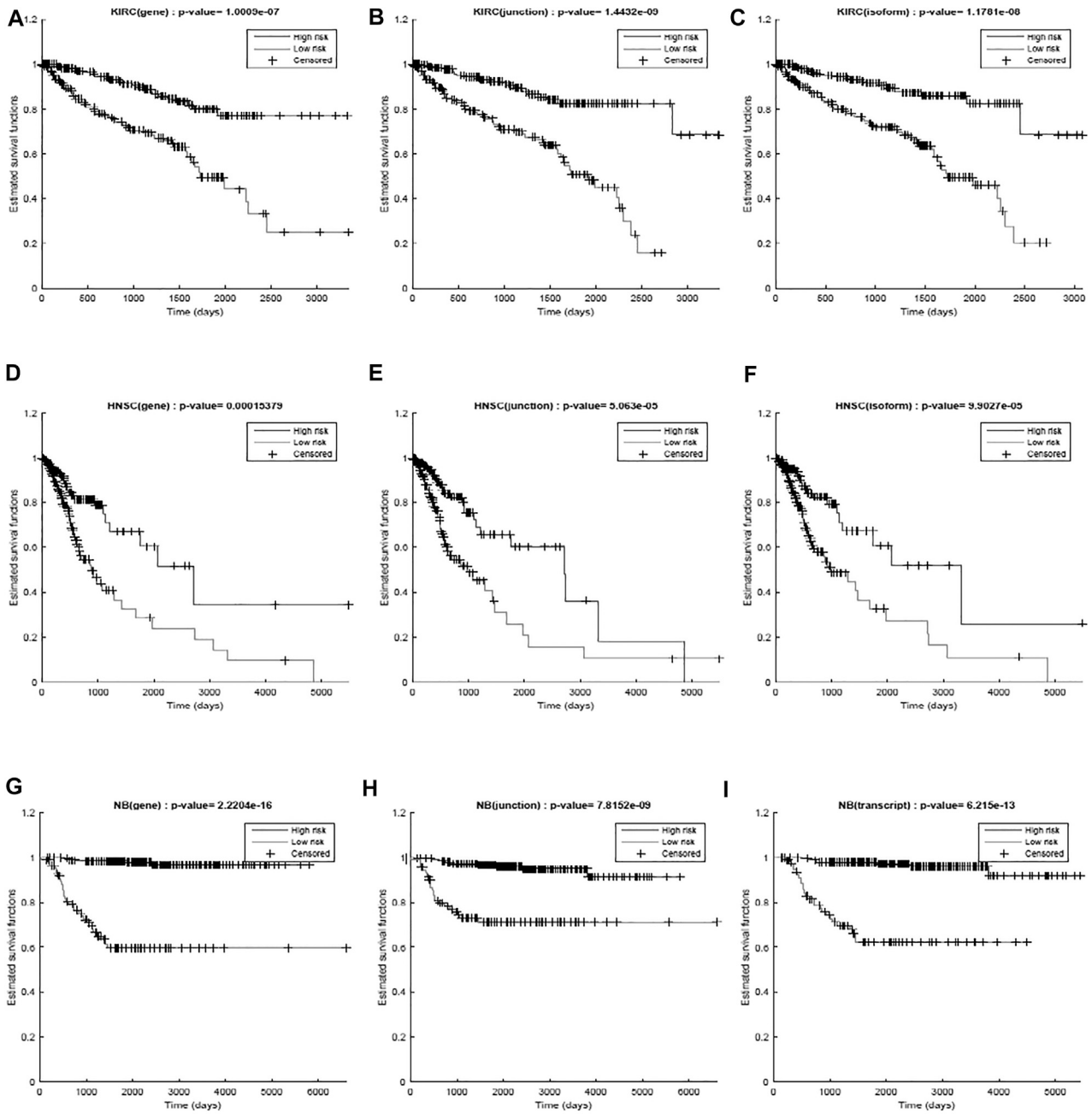

FIGURE 3 | Kaplan–Meier curves and the log-rank *p*-values for the predictions made by the proposed work for different diseases (KIRC, HNSC, and NB) per different modalities (gene, junction, and isoform). **(A)** RNA-seq gene, **(B)** RNA-seq junction, and **(C)** RNA-seq isoform for kidney cancer; **(D)** RNA-seq gene, **(E)** RNA-seq junction, and **(F)** RNA-seq isoform for head and neck cancer; **(G)** RNA-seq gene, **(H)** RNA-seq junction, and **(I)** RNA-seq isoform for neuroblastoma. The curves show a clear separation between the two predicted groups.

overfitting and underfitting problems due to the inappropriate number of selected input features for the baselines, we use a validation set to choose the best number of features output by the mRMR feature selector. The validation sets are also used to tune baselines' hyper-parameters in **Eq. 1**. For the support vector machines, we used the linear kernel as it resulted in the best performance, in which case the performance remains robust with respect to variations in the only model's hyper-parameter, *C*. We performed a grid search to find an appropriate number of input probes, trying all numbers in the range (Belkin et al., 2006; Alipanahi et al., 2015) with increments of five.[3] Furthermore, to make an unbiased comparison between the baselines, for SVM, we

[3]We observed consistent performance drops due to model overfitting if more than 35 probes are selected.

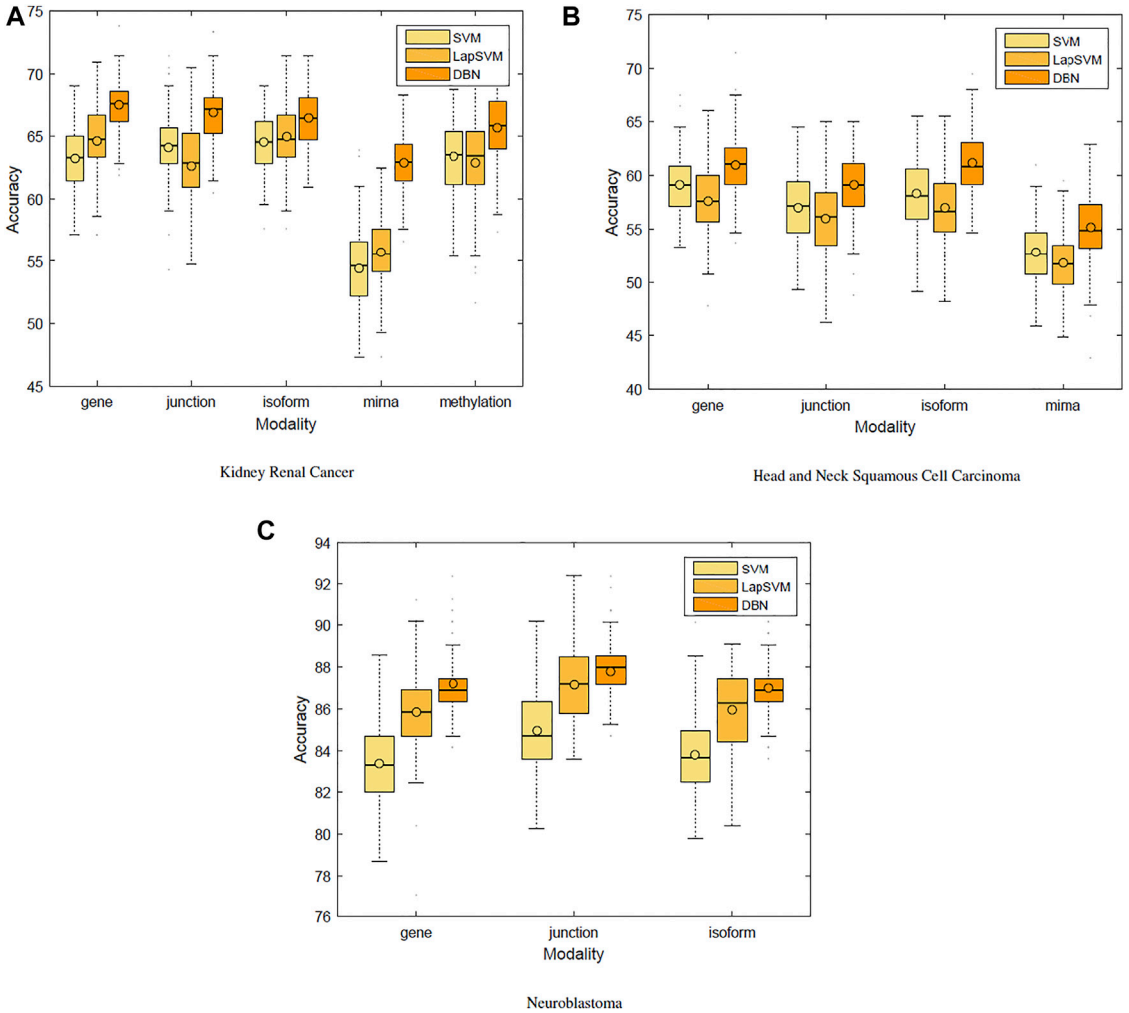

FIGURE 4 | Benchmarking support vector machines, Laplacian SVM, and deep belief networks.

used the same implementation and solver as used in the LapSVM method. For our LapSVM model, on the contrary, we performed a grid search on the set of model's hyper-parameters[4] in addition to the size of selected input probes. We found that the degree of 1, for the Laplacian graph, and the sigma of 3.0 for the model's RBF kernel remain the same across all the sets of parameter configurations. For the extrinsic and intrinsic regularization parameters, we searched the logarithmic search spaces [*1e−2,1e2*] and [*0,1e2*], respectively.

Single-Modality Models

We first used only one modality as input (hence the single-modality model) to show how deep belief networks can be trained on relatively small cancer datasets to predict the survival. **Figure 3** depicts the Kaplan–Meier (KM) curves along with the corresponding log-rank *p*-values for the predictions made by our deep predictor for the three RNA-seq modalities (i.e., gene, junction, and isoform). For each labeled sample, we trained the model once on all but that sample and made a prediction on it, repeated this process for all samples, and plotted the KM curves for the combined predictions. According to the figure, our approach produces meaningful clusters of high-risk and low-risk patients. Furthermore, we benchmarked the proposed predictor against the SVM and the Laplacian SVM (LapSVM). Specifically, we randomly split cancer datasets into the train, test, and validation sets 100 times and subsequently trained deep models and baselines once for each input modality listed in **Table 1**. **Figure 4** illustrates the boxplots of accuracies achieved during this experiment. According to this figure, the trends observed in the prognostic power of individual molecular datasets correlate and strongly depend on the cancer type. Furthermore, the DBN is doing consistently better than baseline methods on average. Importantly, this improvement comes with a tighter confidence interval, as illustrated in **Table 2**. Interestingly, despite the relative success of the semi-supervised LapSVM in leveraging the unlabeled data, it is unable to surpass the supervised SVM for some input

[4]Sigma for the RBF kernel, the number of neighbors, the extrinsic and intrinsic regularization parameters, and the power (degree) of the graph Laplacian to use as the graph regularizer.

TABLE 2 | Mean (SD) of accuracies for 100 randomly initialized runs. The DBN has the smallest variance for the majority of datasets/modalities.

		Modality				
		Gene	**Junction**	**Isoform**	**miRNA**	**Methylation**
Disease	KIRC	63.27 (2.64)	65.13 (2.80)	64.52 (2.53)	54.43 (3.38)	63.39 (3.08)
		64.62 (2.76)	62.65 (3.23)	65.0 (2.99)	55.71 (2.79)	62.88 (3.59)
		67.48 (**2.19**)	66.97 (**2.50**)	66.49 (**2.24**)	62.9 (**2.47**)	65.71 (**3.0**)
	HNSC	59.15 (2.78)	56.94 (3.57)	58.33 (3.12)	54.43 (3.38)	
		57.61 (**3.13**)	55.99 (3.79)	56.96 (3.19)	55.72 (**2.79**)	
		61.04 (3.2)	59.22 (**3.1**)	61.14 (**2.85**)	61.14 (2.85)	
	NB	83.39 (1.88)	84.94 (2.12)	83.81 (1.96)		
		85.86 (2.04)	87.15 (1.86)	85.96 (1.97)		
		87.18 (**1.6**)	87.78 (**1.49**)	86.97 (**1.49**)		

A

	Single modality Accuracy	Accuracy Improvement				
		Gene	Junction	Isoform	miRNA	Methylation
Gene	67.48		-0.27	-0.30	0.28	1.55
Junction	66.88	0.33		0.28	0.08	2.12
Isoform	66.49	0.70	0.66		0.11	2.48
mirna	62.90	4.86	4.05	3.70		2.43
methylation	65.71	3.32	3.29	3.26	-0.38	

B

	Single modality Accuracy	Accuracy Improvement			
		Gene	Junction	Isoform	miRNA
Gene	60.98		0.13	0.58	0.25
Junction	59.16	1.95		2.06	-0.11
Isoform	61.14	0.41	0.08		-0.24
mirna	55.10	6.13	3.95	5.80	

C

	Single modality Accuracy	Accuracy Improvement		
		Gene	Junction	Isoform
Gene	87.21		0.61	0.27
Junction	87.76	0.06		-0.01
Isoform	86.97	0.51	0.78	

FIGURE 5 | Improvement achieved after integrating pairs of modalities. Columns show the additional modality added to the single-modality model (denoted by rows). Cell values show the difference in accuracy between the integrated model and the single-modality model. Tables **(A)–(C)** correspond to KIRC, HNSC, and NB, respectively.

modalities. This is because semi-supervised methods need a significant amount of unlabeled data to learn the distribution of input space efficiently, which does not hold for most cancer datasets. DBN models, however, are more immune to this shortcoming as evidenced by the results presented here.

The Multi-Modality Pipeline

Cancer is known to be a disease associated with dysregulation of cellular mechanisms in different levels. Hence, no single molecular modality is sufficient to predict cancer-related endpoints, such as the survival (Chin and Gray, 2008). Therefore, changes in biological pathways may be captured more accurately if different modalities are integrated together seamlessly. **Figure 4** suggests that the most effective molecular modality regarding the prediction performance is different across different cancer types. For instance, for the kidney cancer dataset, the model trained over the RNA-seq: gene modality results in the most accurate predictor, whereas in case of neuroblastoma, RNA-seq: junction modality leads to the most accurate model. Ideally, we would like to have an integrated pipeline that is more accurate than each of the single modality models individually. Our goal in this section is to examine whether adding another molecular data modality can provide more prognostic power given the proposed integration paradigm through an additional sigmoid layer that is stacked on top of the RBM layer.

Figure 5 shows the heat map of accuracy improvements when pairs of different input modalities are combined according to the final pipeline design. Our results suggest that integration of latent features generated by deep belief networks from different modalities leads to improvements for majority of the cases. This improvement, however, is not significant in case of integration of two RNA-seq modalities. This is because they are different representations of the same source of information and combining them will add little additional predictive value. On the contrary, combining data of different molecular levels can lead to more substantial improvements as is the case with the integration of methylation/miRNA and the RNA-seq data modality. Note that, for the HNSC dataset, miRNA does not provide additional improvement, which is in agreement with recent findings that miRNA is not directly related to the disease prognosis (Hess et al., 2019), as also indicated by its poor prediction accuracy[5] in

[5]The low accuracies reported for HNSC can be attributed to the biological variability of samples, as this cancer type includes cancer of the oral cavity, pharynx, larynx, nasal cavity, and salivary glands.

Figure 4. Hence, its addition does not result in any improvement. Interestingly, despite the large difference in the model's performance when trained on miRNA vs. other profiles, the drop in performance of the integrated model is negligible, suggesting that the integrated model can offer robustness as well as synergistic gains in performance.

CONCLUSION

In this study, we developed a deep learning–based pipeline to predict cancer survival. Because of the unsupervised nature of the pre-training stage, we were able to leverage the unlabeled and censored data to arrive at a better initialization of the model parameters. Such an initialization is a critical step that drives the final fine-tuned model to a more biologically relevant point in the parameter space particularly when the number of layers in the model increases. Our results showed that the proposed model architecture can indeed achieve this goal by successfully exploiting the information that is available in such data and subsequently integrating derived features from different molecular profiles. This is corroborated by the fact that our trained models consistently outperformed the semi-supervised baseline. Moreover, we showed that the most informative data modality can be different across different cancer types, which justifies the need for an integrated decision support system that has the ability to generate synergistic improvements from multiple available modalities. It is worth noting that the focus and scope of this study was on presenting the merits of deep models for extracting informative features from molecular profiles of cancer tumors in an integrated manner. Needless to say that including more modalities such as clinical and proteomic data can enhance the prediction performance even further, as shown in other studies (Liu et al., 2014; Yuan et al., 2014), and can be considered a future work for a more comprehensive decision support system. Another direction that requires further exploration and attention is to evaluate the robustness of such models in light of data scarcity and data variation. The presented approach was an effort to address this challenge by exploiting unlabeled data; however, an important question would be how models trained on data from one study are generalizable and applicable to the data acquired for the same disease but from another study. Finally, it is desirable to know the strengths and limitations of deep belief networks with other pre-training frameworks used for training deep models, such as the variational auto-encoders (An and Cho, 2015) and the more recent contrastive learning (Falcon and Cho, 2020) framework.

Despite all their success in extracting informative latent features from data, deep models are considered black boxes that learn by simple associations and co-occurrences (Mamoshina et al., 2016). This obviates the need for human intervention to generate hand-crafted features or to use the expert knowledge but comes at cost of lacking transparency and interpretability in such models. Making deep interpretable models is currently an active research that has caught attention of researches in the machine learning community and is another dimension where this work can be expanded as a future work.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. These data can be found at https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga

AUTHOR CONTRIBUTIONS

HH, the main author. MW, PI.

REFERENCES

Ahmad, L. G. (2013). Using Three Machine Learning Techniques for Predicting Breast Cancer Recurrence. *J. Health Med. Inform.* 2013, 114. doi:10.1155/2013/290568

Alipanahi, B., Delong, A., Weirauch, M. T., and Frey, B. J. (2015). Predicting the Sequence Specificities of DNA- and RNA-Binding Proteins by Deep Learning. *Nat. Biotechnol.* 33 (8), 831–838. doi:10.1038/nbt.3300

An, J., and Cho, S. (2015). Variational Autoencoder Based Anomaly Detection Using Reconstruction Probability. *Spec. Lecture IE* 2 (1), 1–18. doi:10.5220/0010375905670574

Azizi, S. (2020). "Ultrasound-based Detection of Prostate Cancer Using Automatic Feature Selection with Deep Belief Networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer.

Belkin, M., Niyogi, P., and Sindhwani, V. (2006). Manifold Regularization: A Geometric Framework for Learning from Labeled and Unlabeled Examples. *J. machine Learn. Res.* 7 (Nov), 2399–2434. doi:10.7551/mitpress/2016.003.0012

Bengio, Y. (2007). Greedy Layer-wise Training of Deep Networks. *Adv. Neural Inf. Process. Syst.* 19, 153. doi:10.7551/mitpress/7503.003.0024

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., et al. (2012). The cBio Cancer Genomics portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discovery* 2 (5), 401–404. doi:10.1158/2159-8290.CD-12-0095

Chapelle, O., Scholkopf, B., and Zien, Eds., A. (2009). Semi-Supervised Learning (Chapelle, O. et al., Eds.; 2006) [Book reviews]. *IEEE Trans. Neural Netw.* 20 (3), 542. doi:10.1109/tnn.2009.2015974

Chin, L., and Gray, J. W. (2008). Translating Insights from the Cancer Genome into Clinical Practice. *Nature* 452 (7187), 553–563. doi:10.1038/nature06914

Collisson, E. A., Bailey, P., Chang, D. K., and Biankin, A. V. (2019). Molecular Subtypes of Pancreatic Cancer. *Nat. Rev. Gastroenterol. Hepatol.* 16, 207–220. doi:10.1038/s41575-019-0109-y

Ditzler, G., Polikar, R., and Rosen, G. (2015). Multi-Layer and Recursive Neural Networks for Metagenomic Classification. *IEEE Trans.on Nanobioscience* 14 (6), 608–616. doi:10.1109/tnb.2015.2461219

Erhan, D. (2010). Why Does Unsupervised Pretraining Help Deep Learning?. *J. Machine Learn. Res.* 11 (Feb), 625–660. doi:10.1007/978-3-030-11479-4_13

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., et al. (2017). Dermatologist-level Classification of Skin Cancer with Deep Neural Networks. *Nature* 542 (7639), 115–118. doi:10.1038/nature21056

Fakoor, R. (2020). "Using Deep Learning to Enhance Cancer Diagnosis and Classification," in Proceedings of the International Conference on Machine Learning.

Falcon, W., and Cho, K. (2020). A Framework for Contrastive Self-Supervised Learning and Designing a New Approach. arXiv preprint arXiv:2009.00104

Fan, X.-N., and Zhang, S.-W. (2015). lncRNA-MFDL: Identification of Human Long Noncoding RNAs by Fusing Multiple Features and Using Deep Learning. *Mol. Biosyst.* 11 (3), 892–897. doi:10.1039/c4mb00650j

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning (Adaptive Computation and Machine Learning Series)*. Massachusetts: MIT Press.

Hassanzadeh, H. R., Sha, Y., and Wang, M. D. (2017). "DeepDeath: Learning to Predict the Underlying Cause of Death with Big Data," in Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE. doi:10.1109/embc.2017.8037579

Hassanzadeh, H. R., and Wang, M. D. (2016). "DeeperBind: Enhancing Prediction of Sequence Specificities of DNA Binding Proteins," in Bioinformatics and Biomedicine (BIBM). IEEE International Conference onIEEE. doi:10.1109/bibm.2016.7822515

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *Unsupervised Learning. The Elements of Statistical Learning*. Berlin: Springer, 485–585. doi:10.1007/978-0-387-84858-7_14

Hess, J., Unger, K., Maihoefer, C., Schüttrumpf, L., Wintergerst, L., Heider, T., et al. (2019). A Five-microRNA Signature Predicts Survival and Disease Control of Patients with Head and Neck Cancer Negative for HPV Infection. *Clin. Cancer Res.* 25 (5), 1505–1516. doi:10.1158/1078-0432.ccr-18-0776

Hinton, G. E., and Salakhutdinov, R. R. (2006). Reducing the Dimensionality of Data with Neural Networks. *Science* 313 (5786), 504–507. doi:10.1126/science.1127647

Hsieh, J. J., Le, V., Cao, D., Cheng, E. H., and Creighton, C. J. (2018). Genomic Classifications of Renal Cell Carcinoma: a Critical Step towards the Future Application of Personalized Kidney Cancer Care with Pan-Omics Precision. *J. Pathol.* 244 (5), 525–537. doi:10.1002/path.5022

Kim, D., Shin, H., Song, Y. S., and Kim, J. H. (2012). Synergistic Effect of Different Levels of Genomic Data for Cancer Clinical Outcome Prediction. *J. Biomed. Inform.* 45 (6), 1191–1198. doi:10.1016/j.jbi.2012.07.008

Kim, J., and Shin, H. (2013). Breast Cancer Survivability Prediction Using Labeled, Unlabeled, and Pseudo-labeled Patient Data. *J. Am. Med. Inform. Assoc.* 20 (4), 613–618. doi:10.1136/amiajnl-2012-001570

Kim, W., Kim, K. S., Lee, J. E., Noh, D.-Y., Kim, S.-W., Jung, Y. S., et al. (2012). Development of Novel Breast Cancer Recurrence Prediction Model Using Support Vector Machine. *J. Breast Cancer* 15 (2), 230–238. doi:10.4048/jbc.2012.15.2.230

Kumar, D., Wong, A., and Clausi, D. A. (2015). "Lung Nodule Classification Using Deep Features in Ct Images," in 12th Conference on Computer and Robot Vision (CRV). doi:10.1109/crv.2015.25

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep Learning. *Nature* 521 (7553), 436–444. doi:10.1038/nature14539

Li, B., and Dewey, C. N. (2011). RSEM: Accurate Transcript Quantification from RNA-Seq Data with or without a Reference Genome. *BMC bioinformatics* 12 (1), 323. doi:10.1186/1471-2105-12-323

Liang, M., Li, Z., Chen, T., and Zeng, J. (2015). Integrative Data Analysis of Multi-Platform Cancer Data with a Multimodal Deep Learning Approach. *Ieee/acm Trans. Comput. Biol. Bioinf.* 12 (4), 928–937. doi:10.1109/tcbb.2014.2377729

Liu, Z., Zhang, X.-S., and Zhang, S. (2014). Breast Tumor Subgroups Reveal Diverse Clinical Prognostic Power. *Scientific Rep.* 4, 4002. doi:10.1038/srep04002

Mamoshina, P., Vieira, A., Putin, E., and Zhavoronkov, A. (2016). Applications of Deep Learning in Biomedicine. *Mol. Pharmaceutics* 13 (5), 1445–1454. doi:10.1021/acs.molpharmaceut.5b00982

Melacci, S., and Belkin, M. (2011). Laplacian Support Vector Machines Trained in the Primal. *J. Machine Learn. Res.* 12 (Mar), 1149–1184. doi:10.7551/mitpress/7496.003.0004

Miotto, R., Wang, F., Wang, S., Jiang, X., and Dudley, J. T. (2017). Deep Learning for Healthcare: Review, Opportunities and Challenges. *Brief. Bioinform.* 19 (6), 1236–1246. doi:10.1093/bib/bbx044

Park, K., Ali, A., Kim, D., An, Y., Kim, M., and Shin, H. (2013). Robust Predictive Model for Evaluating Breast Cancer Survivability. *Eng. Appl. Artif. Intelligence* 26 (9), 2194–2205. doi:10.1016/j.engappai.2013.06.013

Park, Y., and Kellis, M. (2015). Deep Learning for Regulatory Genomics. *Nat. Biotechnol.* 33, 825–826. doi:10.1038/nbt.3313

Peng, H., Long, F., and Ding, C. (2005). Feature Selection Based on Mutual Information: Criteria of max-dependency, max-relevance, and Min-Redundancy. *IEEE Trans. Pattern Anal. Mach Intell.* 27 (8), 1226–1238. doi:10.1109/TPAMI.2005.159

Sicklick, J. K., Kato, S., Okamura, R., Schwaederle, M., Hahn, M. E., Williams, C. B., et al. (2019). Molecular Profiling of Cancer Patients Enables Personalized Combination Therapy: the I-PREDICT Study. *Nat. Med.* 25 (5), 744–750. doi:10.1038/s41591-019-0407-5

Spencer, M., Eickholt, J., and Cheng, J. (2015). A Deep Learning Network Approach to Ab Initio Protein Secondary Structure Prediction. *Ieee/acm Trans. Comput. Biol. Bioinf.* 12 (1), 103–112. doi:10.1109/tcbb.2014.2343960

Thierry-Mieg, D., and Thierry-Mieg, J. (2006). AceView: a Comprehensive cDNA-Supported Gene and Transcripts Annotation. *Genome Biol.* 7 (1), S12. doi:10.1186/gb-2006-7-s1-s12

Tseng, C.-J., Lu, C.-J., Chang, C.-C., and Chen, G.-D. (2014). Application of Machine Learning to Predict the Recurrence-Proneness for Cervical Cancer. *Neural Comput. Applic* 24 (6), 1311–1316. doi:10.1007/s00521-013-1359-1

Wang, S., Peng, J., Ma, J., and Xu, J. (2016). Protein Secondary Structure Prediction Using Deep Convolutional Neural fields. *Scientific Rep.* 6, 121. doi:10.1038/srep18962

Wang, Z., Jensen, M. A., and Zenklusen, J. C. (2016). *A Practical Guide to the Cancer Genome Atlas (TCGA). Statistical Genomics*. Berlin: Springer, 111–141. doi:10.1007/978-1-4939-3578-9_6

Weinstein, J. N., au, fnm., Collisson, E. A., Mills, G. B., Shaw, K. R. M., Ozenberger, B. A., et al. (2013). The Cancer Genome Atlas Pan-Cancer Analysis Project. *Nat. Genet.* 45 (10), 1113–1120. doi:10.1038/ng.2764

Yuan, Y., Van Allen, E. M., Omberg, L., Wagle, N., Amin-Mansour, A., Sokolov, A., et al. (2014). Assessing the Clinical Utility of Cancer Genomic and Proteomic Data across Tumor Types. *Nat. Biotechnol.* 32 (7), 644–652. doi:10.1038/nbt.2940

Zeng, J., Cai, H., and Akutsu, T. (2020). "Breast Cancer Subtype by Imbalanced Omics Data through A Deep Learning Fusion ModelBiochem. Bioinformatics," in Proceedings of the 2020 10th International Conference on Bioscience. doi:10.1145/3386052.3386063

Zhang, W., Yu, Y., Hertwig, F., Thierry-Mieg, J., Zhang, W., Thierry-Mieg, D., et al. (2015). Comparison of RNA-Seq and Microarray-Based Models for Clinical Endpoint Prediction. *Genome Biol.* 16 (1), 133–141. doi:10.1186/s13059-015-0694-1

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Big Data

ORIGINAL RESEARCH
published: 27 August 2021
doi: 10.3389/fdata.2021.689358

Computer Vision for Continuous Bedside Pharmacological Data Extraction: A Novel Application of Artificial Intelligence for Clinical Data Recording and Biomedical Research

Logan Froese[1]†‡, Joshua Dian[2]†‡, Carleen Batson[3]†, Alwyn Gomez[2,3]†, Amanjyot Singh Sainbhi[1]†, Bertram Unger[4]† and Frederick A. Zeiler[1,2,3,5,6]†*

[1]Biomedical Engineering, Faculty of Engineering, University of Manitoba, Winnipeg, MB, Canada, [2]Section of Neurosurgery, Department of Surgery, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, MB, Canada, [3]Department of Anatomy and Cell Science, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, MB, Canada, [4]Section of Critical Care, Department of Medicine, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, MB, Canada, [5]Centre on Aging, University of Manitoba, Winnipeg, MB, Canada, [6]Division of Anaesthesia, Department of Medicine, Addenbrooke's Hospital, University of Cambridge, Cambridge, United Kingdom

Introduction: As real time data processing is integrated with medical care for traumatic brain injury (TBI) patients, there is a requirement for devices to have digital output. However, there are still many devices that fail to have the required hardware to export real time data into an acceptable digital format or in a continuously updating manner. This is particularly the case for many intravenous pumps and older technological systems. Such accurate and digital real time data integration within TBI care and other fields is critical as we move towards digitizing healthcare information and integrating clinical data streams to improve bedside care. We propose to address this gap in technology by building a system that employs Optical Character Recognition through computer vision, using real time images from a pump monitor to extract the desired real time information.

Methods: Using freely available software and readily available technology, we built a script that extracts real time images from a medication pump and then processes them using Optical Character Recognition to create digital text from the image. This text was then transferred to an ICM + real-time monitoring software in parallel with other retrieved physiological data.

Results: The prototype that was built works effectively for our device, with source code openly available to interested end-users. However, future work is required for a more universal application of such a system.

Conclusion: Advances here can improve medical information collection in the clinical environment, eliminating human error with bedside charting, and aid in data integration for biomedical research where many complex data sets can be seamlessly integrated digitally. Our design demonstrates a simple adaptation of current technology to help with this integration.

Keywords: computer vision, image modification, opitcal character recognition, system integration, data integration

OPEN ACCESS

Edited by:
Bernard Kamsu Foguem, Université de Toulouse, France

Reviewed by:
Shivanand Sharanappa Gornale, Rani Channamma University, India
Jatinderkumar Saini, Symbiosis Institute of Computer Studies and Research (SICSR), India

****Correspondence:***
Logan Froese
log.froese@gmail.com

†ORCID:
Logan Froese
orcid.org/0000-0002-6076-0189
Joshua Dian
orcid.org/0000-0002-2193-4916
Carleen Batson
orcid.org/0000-0002-7928-8523
Alwyn Gomez
orcid.org/0000-0002-3737-2065
Amanjyot Singh Sainbhi
orcid.org/0000-0003-3231-5683
Bertram Unger
orcid.org/0000-0002-5739-3955
Frederick A. Zeiler
orcid.org/0000-0003-1737-0510

‡These authors share first authorship

Specialty section:
This article was submitted to Medicine and Public Health, a section of the journal Frontiers in Big Data

Received: *09 April 2021*
Accepted: *09 August 2021*
Published: *27 August 2021*

Citation:
Froese L, Dian J, Batson C, Gomez A, Sainbhi AS, Unger B and Zeiler FA (2021) Computer Vision for Continuous Bedside Pharmacological Data Extraction: A Novel Application of Artificial Intelligence for Clinical Data Recording and Biomedical Research. Front. Big Data 4:689358. doi: 10.3389/fdata.2021.689358

INTRODUCTION

Current therapeutic interventions in Traumatic Brain Injury (TBI) are generally based on low frequency physiological response over large sample sizes, focusing on long epoch outcomes (Chalmers et al., 1981; Carney et al., 2017). Though this methodology can be effective in identifying large global phenomenon, momentary individualized events are masked within these large datasets. Thus, methodologies are emerging that leverage higher frequency data to find momentary phenomenon that focus on individualized patient response to medical treatment (Carney et al., 2017; Matchett et al., 2017; Zeiler et al., 2018a). Furthermore, within TBI care, recent literature has emerged connecting high frequency physiology with TBI outcome (Balestreri et al., 2015; Cabella et al., 2017; Zeiler et al., 2018b). Yet, few studies connect the momentary response of high frequency physiology to current hourly recorded therapeutic infusions (Froese et al., 2020a; Froese et al., 2020b; Klein et al., 2020). Through the use of more robust and individualized datasets, treatment guidelines can be focused on patient specific healthcare interventions which can lead to more individualized and personalized care. To take advantage of emerging technologies and new health metrics, real time high frequency physiological and treatment care data needs to be recorded and integrated. However, despite this increase in computational integration within health care, there are countless devices that are either released with insufficient digital output or are simply too outdated to carry the necessary hardware infrastructure to output the required data at a high frequency. This is particularly the case with many commercially available and clinically utilized medication pumps. As such, treatment information in many instances is still recorded manually at low frequency in bedside charts, or e-charts. Such methods are prone to errors in data entry and are time consuming for clinical staff.

The limited compatibility of many bedside medical devices hinders clinicians' ability to capture high frequency data, thus there is a need to leverage interfaces that convert such data from bedside devices directly into digital data. Many medical devices use text displays to convey the required information to the user. The text display therefore has the desired information, but based on the antiquated hardware, it lacks the compatibility to convert the information to a digital format. This problem is described as Text Information Extraction (TIE) (Jung et al., 2004) and has been addressed in other environments like text-based image processing, (Park et al., 1999; Kim et al., 2002; Carvalho, 2016) document decoding (Cheng et al., 1997; Feng et al., 2006) and video text extraction (Locating Characters in Sc, 1047; Fischer et al., 1995). All of these systems extract alphanumeric characters using Optical Character Recognition (OCR) via computer vision techniques, which leverage artificial intelligence to convert image characters into digital data (Schantz, 1982). This method, although well documented, has yet to be adapted for the use and conversion of medical monitoring equipment. Therefore, with the emergence of new openly available software and the universal nature of personal computers, there is a potential to adapt past medical devices to the computational age.

Furthermore, for the integration of many older medical devices the only feasible solution to digital integration is through the use of scripting (Carvalho, 2013; Delaney et al., 2013; Carvalho, 2021). Likewise, as clinical data collection exceeds the limits of humans, the need to leverage scripting to ensure accurate data collection becomes necessary (Mardis, 2011; Delaney et al., 2013). To bridge this gap in compatibility, we have endeavoured to build a system that uses a camera to attain real time output from a text based display screen from bedside intravenous medication pumps and convert it into a continuously updated digital format to be captured and linked with other time-series data at the bedside in real time.

MATERIALS AND METHODS

Device Set-Up and Image Capture

This work was conducted at the Winnipeg Acute TBI Laboratories, at the University of Manitoba. The set-up consisted of a USB connected camera (Logitech C920s Pro HD Webcam, Logitech, Newark, CA, United States) to take real time images of a commercially and commonly available intravenous medication pump (Baxter Colleague 3 CXE, Baxter Canada, Mississauga, Canada) which currently has no digital outport. Images are captured at 60 frames/second from a USB camera and copied directly onto a basic consumer laptop, see **Figure 1**. The full Python scripting language code (Python 3, Scotts Valley, CA: CreateSpace) can be found in either **Supplementary Appendix A** or GitHub (https://github.com/lofro/TIE_OCR). The basic operation of this system leverages 4 main libraries in python; "pytesseract," "cv2," "serial" and "tkinter." "pytesseract" and is used for the OCR processing. (Lee, 2007) "Cv2" is also an image processing and manipulation library. (Bradski, 2000) The use of these libraries will be detailed in the subsections to follow. "Serial" is a library in python that allows for the creation and use of serial sockets (Welcome to PySerial's Documentation PySerial 3.4 Documentation). Finally, we used the "tkinter" library to create the display and user interface that is seen in **Figure 2** (Lundh, 1999). To create a video we leveraged the "cv2. CaptureVideo" function to extract frames and the "tkinter. Canvas" to display these frames. When either the *snapshot* button is press or the time delay is reached, the current frame captured will be processed.

Image Processing and Feature Extraction

The TIE for these images was performed using Python. On the initiation of the code, an interface for the image capture will appear, as shown in **Figure 2**. The subsequent image manipulations are demonstrated in **Figure 3**, which illustrates our method to solve the TIE problem. The TIE problem can be divided into the following sub-problems: detection, localization, tracking, extraction/enhancement, and recognition (Jung et al., 2004). Within our design we focused on localization, extraction/enhancement and recognition, as we can assume the images

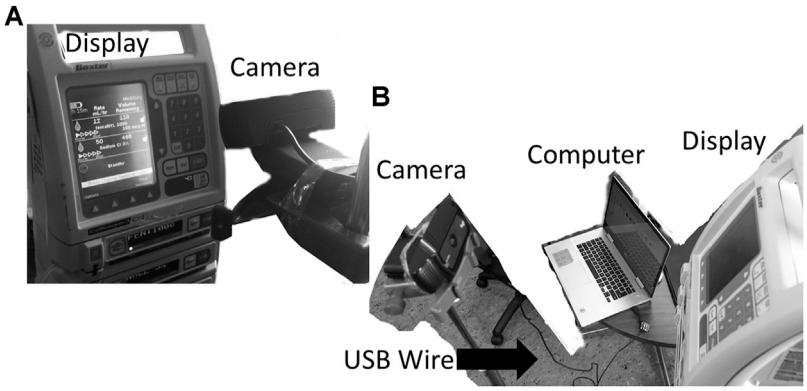

FIGURE 1 | Setup for the camera and pump. General setup for our design, with the monitor display being captured through an external camera is displayed in image **(A)** and **(B)**. In figure **B** the USB wire connecting the computer to camera can be seen. The current design has the camera directly in front of the text display.

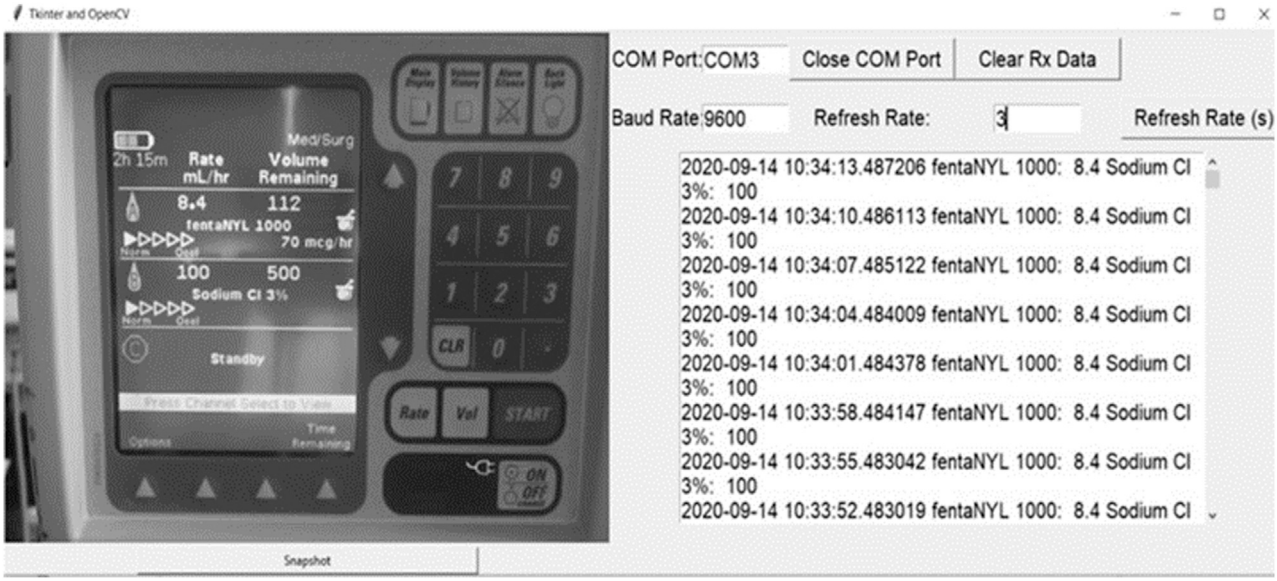

FIGURE 2 | Python interface. **Left Panel**–Displays digital photo of medication pump taken by the camera, **Right Panel** - Displays the interface of our system, with real time data being updated from the extracted features from the medication pump display.

captured have some form of desired information, and that the features of interest stay relatively constant.

An image can be captured manually or automatically after an allotted time. Once the image is captured it goes through the entire TIE image processing as seen in **Figure 3**, proceeding alphabetically going from A to H. The two TIE subgroups of image localization and extraction/enhancement are performed in unison, shown in **Figures 3A–E**. Initially the image is converted to grayscale using a "cv2" function (**Figure 3A**), then using Canny edge detection, the image edges are traced (**Figure 3B**). (Open(X). Canny Edge) Canny highlights the edges of an image using the intensity gradient of the image, which is the color difference on local pixels to find the edge of shapes within the image (Canny, 1986). Using these edges, we can differentiate the display from the larger image by the rectangular aspect of the display. To do this the edges are grouped into contours. Contours are the bounding points that map the outline of a continuous white shape of **Figure 3B**. Each continuous white shape is bounded by the smallest, best fitting rectangle that contains all the contours of that group. With all shapes having a respective bounding rectangle, the largest area rectangle can be found, which is assumed to be the display screen and used to give **Figure 3C**.

The image is then enlarged to improve the small feature edges for the adaptive mean threshold. The adaptive mean threshold uses the area of local pixel brightness to find a mean brightness which then can be contrasted against the pixel of interest to identify if it should be black or white, resulting in **Figure 3D**. (Open(X). Image Thres) Next, the contours of **Figure 3D** are

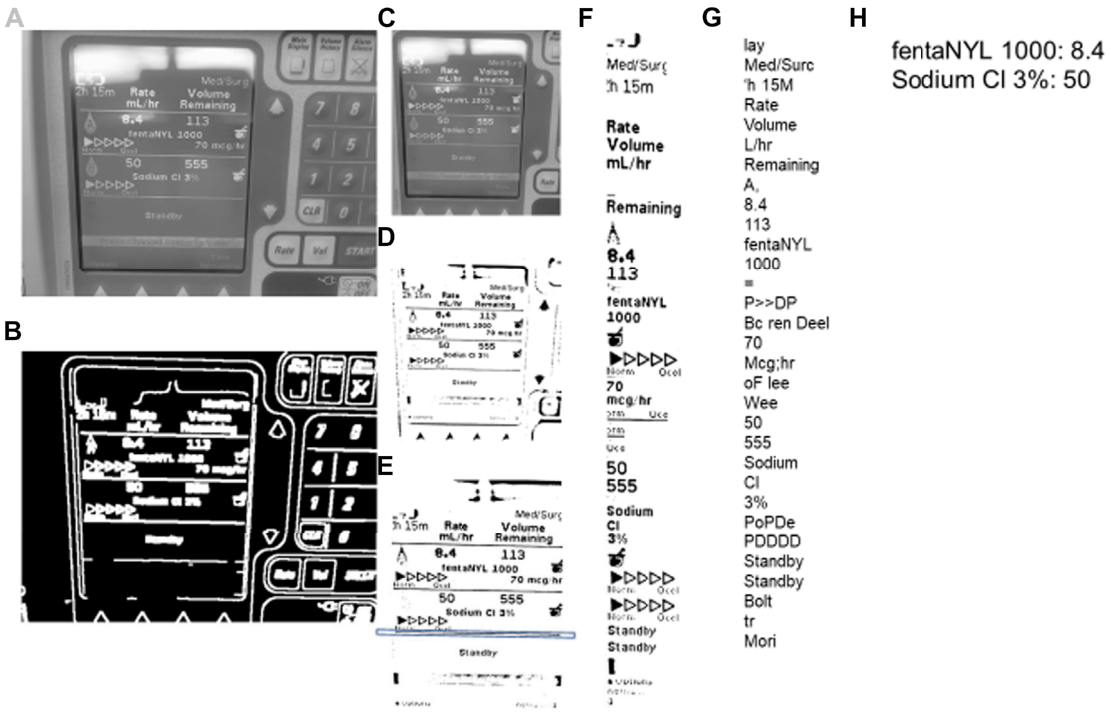

FIGURE 3 | Steps for image processing–TIE and OCR. TIE = Text Information Extraction, OCR = Optical Character Recognition. The processing proceeds in alphabetical order. **(A)** is the initial image converted to grayscale. **(B)** is the grayscaled image processed with the Canny function. **(C)** is the grayscaled cropped image using the rectangle contours of image **(B)**. **(D)** is the adaptive mean threshold function of image **(C)**. E is the cropped and rotated image **(D)** with the key horizontal line contained within a box. F is the features found with the google tesseract of image **(E)**. **(G)** is the string that Google Tesseract output from image **(F)**. **(H)** is the final output of the **(G)** string process.

found in a similar way as before, (using canny edge detection) with the exception that it looks for the continuous black portions. Like before, the continuous black shapes are all bounded by a rectangle and used with their respective contours to rotate the image and crop the image for a second time. To rotate the image, a key horizontal line is needed (highlighted by the box around a line in **Figure 3E**), this line is found by using the relative height to length of the bounding rectangle. The bounding rectangle must have a width greater than ¾ of the image width, and of the rectangles that meet this criterion, the one with the smallest height is chosen. Next, with the contours from which the previously described bounding rectangle encompasses, the line of best fit is made. That being, a best fit line is drawn through the key horizontal line. This is the least squares regression line with the contours as the points of interest. The best fit line is created using a "cv2" function and has an output of a location and an angle of rotation. (Bradski, 2000) This angle of rotation is also the angle for the image to rotate. To find the cropping area, the width and location of the bounding rectangle for the key horizontal line is used to find the x component of the cropped image (the horizontal location and width). The y component (vertical location and height) is assumed to be at the 5 and 90% of the initial image height, which allows the image to be cropped (**Figure 3E**). This concludes the localization of the TIE process as the image is focused on only the text display. The last step in enhancement/extraction is performed using Google Tesseract's (Google Inc., https://github.com/tesseract-ocr/tesseract/) feature selection function, this function uses an artificial intelligence algorithm to find all key shapes within the image. (Lee, 2007) These are then cropped from the initial image and displayed in a consecutive order to give **Figure 3F**.

Character Recognition

The last part of the TIE process, recognition, uses Google Tesseract OCR (Lee, 2007) to give the output text shown in **Figure 3G**. This process, like all OCR, involves comparing a library of identified shapes to the data, in this way the best matched letter is assumed. (Lee, 2007) From **Figure 3G** the desired values are extracted based on the nature of the OCR output and design of the text display, that being, the dose is always followed by the dose amount and left/time remaining, and the medication type is found by a list of predefined words of interest. Together the dose amount and medication can be paired up, and in almost any fashion given as **Figure 3H**. To improve accuracy, we found the key words (those being greater then 4 characters of alphabetical values) and connected those with a number in a similar location, for the full OCR code see **Supplementary Appendix A.2**. From here the data is digitized and can be output into any desired format. A full process map of the above TIE and OCR processes, from image capture to serial output can be seen in **Figure 4**.

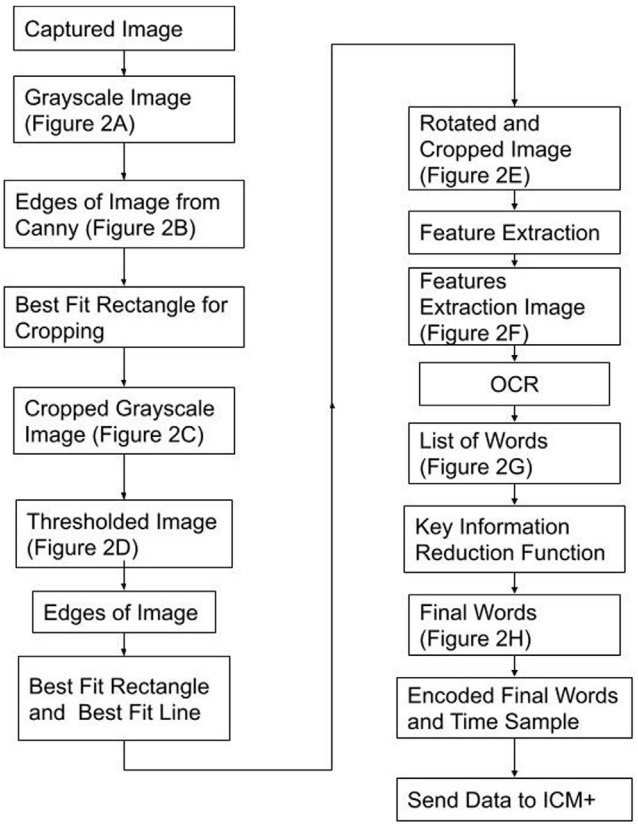

FIGURE 4 | Process Map–From Image Capture to Serial Output Figure displays the process taken to convert the image into its digital information and the steps to send the data to ICM+ (data acquisition platform). The best fit rectangles and line are the key shapes used to crop and rotate the image. Feature Extraction is a Google Tesseract function to find key shapes. Key Information Reduction Function is a function used to find and order the alphanumeric of interest.

Digitized Data Capture

Using a virtual serial port, we sent the serialized data (**Figure 3H**) to Intensive Care Monitoring "Plus" (ICM+) (Cambridge Enterprise Ltd., Cambridge, United Kingdom, http://icmplus.neurosurg.cam.ac.uk), generating continuously updating real time data (**Figure 2**). The virtual serial port is an internal design that acts like serial port for any RS232 ASCII streaming device and was made using freely available software (null-modem emulator (com0com), http://com0com.sourceforge.net). (Hatchett, 1991) In ICM+ the data was parsed into the desired functions identical to the parsing of any other device data. ICM+ was used as an example of a data acquisition platform for the continuous time-series capture of such data, as it is the platform utilized by our laboratory for bedside physiology research. The above-described design can be integrated with any data acquisition platform which can record serial data.

Finally, to show-case the capture of continuous medical pump data in conjunction with other monitoring devices, we recorded continuous bifrontal cerebral regional oxygen saturations using near infrared spectroscopy (Covidien INVOS 7100, Medtronic Canada) and continuous non-invasive arterial blood pressure through a finger-cuff technique (Finapres NOVA Nanocare, Finapres Medical Systems, Enschede, Netherlands, http://www.finapres.com/home), in a volunteer. The regional oxygen saturation was sampled at 1 Hz, while the arterial blood pressure was sampled at 250 Hz. Therefore, we can run our system in parallel with any number of compatible devises as can be seen in **Figure 5**.

RESULTS AND DISCUSSION

System Performance

As this entire system was a proof of concept, the design proves that there is technology available to complete an effective TIE process on a human-based text interface output, using an intravenous medication pump as an exemplar (examples of captured frames that worked can be seen in **Supplementary Appendix B**). Furthermore, the design used only a common camera, a laptop and freely available open source software, (Lee, 2007; Bradski, 2000; Hatchett, 1991) demonstrating the accessibility of this conversion system.

Though we built a working prototype, there were some key issues that arose when operating the system. The first and perhaps most important, is the slightly inconsistent nature of the OCR recognition which has been documented in the past (Carvalho, 2016; Schantz, 1982; Lee, 2007). When implementing OCR, there is a tendency for letters and word orders to be mismatched. For example, a common error is the letter "f" interpreted as a "t," i.e. "tentanyl" instead of "fentanyl." This can be bypassed by backend language algorithms and deep learning techniques (Mokhtar et al., 2018; Le et al., 2019). Another common issue encountered is the mismatch of numbers "5," "6", "8" and "9," which in operation have become interchangeable with one another if the image is insufficiently processed. To overcome this problem in operation, converting the image to **Figure 3F**, with significant space between the lines of text, improved recognition. Also, the enlargement of features made the edges more robust (improving extraction/enhancement of the image). Though it must be acknowledged, in our described design and camera setup, we did not require these improvements to get sufficiently accurate data. Such modifications may be necessary with cheaper and lower resolution cameras.

The second issue is the interference that background noise can have on the image, which interferes with extraction and enhancement. If the display is dim, with a light that reflects directly in the camera, there are scenarios in which the captured image data can be masked behind this light. Likewise, if the camera is moved into such an angle as to obscure the image, the OCR software fails to accurately extract the information. Currently, there are no working examples that we know of that effectively adjust images at obscure angles to effectively output a coherent final image however, there are emerging proposed solutions (Oakley and Satherley, 1998; Li and Doermann, 1999; Li et al., 2000; Witten et al., 2004). Therefore, in the implementation of this design the most

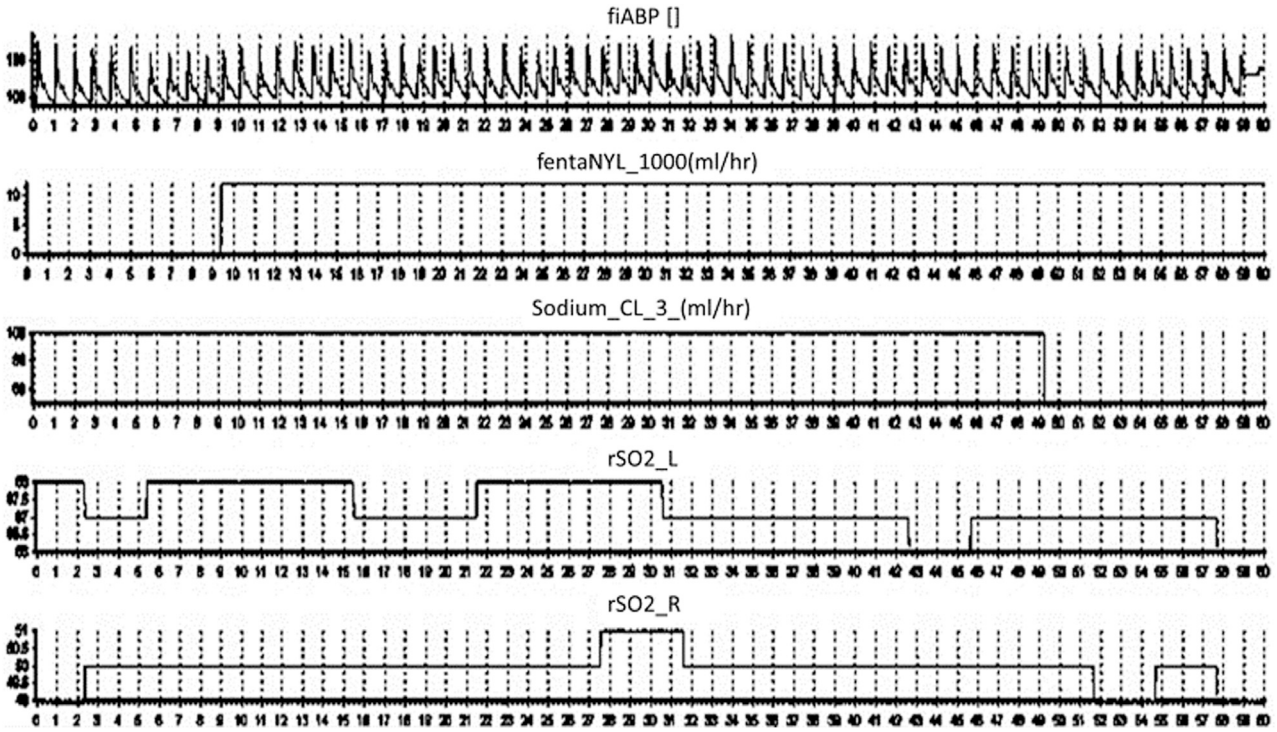

FIGURE 5 | ICM + Final Output. Displays the final output on ICM + over a 60 s period from top to bottom; the arterial blood pressure, fentanyl, sodium chloride and regional oxygen saturations.

effective solution is setting up the camera to extract clear, centered images.

Reflections on Impact of the Designed System

The TIE/OCR Process

This system of converting real time data from a medical device display into digital data, is the first that we have knowledge of. As such, this system illustrates that there is a bridge between computers and older devices that lack the necessary compatibility, using TIE processing. In this way there is an opportunity to extract data even when there is no capability of directly accessing the digital port, or when no digital output is offered. However, the design and operation of this system enforces the desire for a robust TIE methodology, due to the tenuous precision in the output. The mixed precision is caused by errors mostly relating to the OCR methodology for recognition, thus the field of text extraction is expanding with new developments and emerging improvements to all aspects of the TIE processing. These include word detection using Markov Random Field (Yalniz and Manmatha, 2019) and canonical correlation analysis, enhancing image quality by layering multiple images, (Wemhoener et al., 2013) smoothing edges by using corner detection, (Yalniz and Manmatha, 2012) and having more robust feature detection methods (Witten et al., 2004; Oakley and Satherley, 1998; Li et al., 2000; Li and Doermann, 1999) with more areas and designs proposed to improve information retrieval from images (Allan et al., 2003). These improvements highlight ideas to incrementally change the TIE methodology and enhance text extraction. Furthermore, by leveraging Deep Learning techniques before and after the OCR process, the shortcomings that are inherit with the OCR could be addressed. The two key areas to apply these Deep Learning solutions would be the creation of the improved text images (**Figures 3E,F**) and error correction (**Figures 3G,H**), which have emerging methods to address them (Mokhtar et al., 2018; Le et al., 2019; Namysl and Konya, 2019; Yin et al., 2019; Karthikeyan et al., 2021).

For individuals who endeavour to build a similar TIE system, the use of a prebuilt OCR is recommended. The open-source nature of Google Tesseract OCR makes it easily adaptable but supported under the Google banner also gives it access to a vast database to build its character recognition library on. As well, Google Tesseract OCR offers language conversion for over 50 different languages. (Lee, 2007) As global health becomes integrated, systems that can be adapted for a global community become imperative. These platforms bear the added benefit of being supported by a wide group of people, improving not only its functionality but its robustness as it pertains to various aspects including varying text font styles and languages. Therefore, although in theory it is possible to build one's own OCR system, there is limited practical reason to do so.

Application to Bedside Medical Big Data

Aside from the novel application of computer vision to solve a digitization problem for medical device data, the TIE also offers the removal of the human element within data collection, as humans account for a large amount of the inconsistency within data processing (Barchard and Pace, 2011). In both the clinical care provision and biomedical research fields, data accuracy is critical. Errors in bedside or e-chart data entry, associated with human-based methods, can impact care delivery and safety for patients by allowing for treatment decisions to be made on inaccurate information. Similarly, accuracy of data in biomedical research is paramount as the focus of care becomes more responsive and individualized.

The TIE also improves the volume and frequency of data collection from such medical devices, exponentially higher than any human-based recording method. In almost all clinical data extraction, but in particular TBI data, the treatment methodologies are often updated at an hourly rate, with limited concern for the minute-to-minute fluctuations within care. Emerging studies in TBI research identify an optimal cerebral perfusion pressure which is coupled to minute to minute changes in physiology, (Steiner et al., 2002; Aries et al., 2012), with measures like intracranial pressure being well documented as having targeted goals to achieve (Cabella et al., 2017; Carney et al., 2017; Zeiler et al., 2020). Such targets require the implementation of high frequency data analysis, however the treatments associated with these goals is either undocumented, or lack precision in documentation as to the exact momentary changes within care. Thus, methods to improve time resolution, allowing data to be linked with other physiologic information for a clearer picture of treatment response/effect, is required, as highlighted in our example in **Figure 5**. Moreover, the nature of digitized information makes the update, dissemination, and archiving to prevent data loss a nearly trivial task. Thus, the breakdown or damage to one device can be mitigated by having continuous multi-connected data streams, limiting data loss.

FUTURE DIRECTIONS

Despite the novel and interesting results described, future work is required in this area for further optimization. For this type of design there is a need to focus on three basic future implementations: the first, is creating a more robust TIE process with a focus on image enhancement and recognition. Such work will encompass variation in camera face angles and screen brightness/hues. The goal is to improve the efficiency of the output to more suitably honed results. Thus, the implementation of some previously proposed solutions to the OCR process using Deep Learning methods will be explored, including; convolutional neural networks, (Allan et al., 2003) neural machine translation techniques (Mokhtar et al., 2018) and provide improved lexicons.

The second area to address is a refined layout and interface. The goal for this style of technology is to have any user intuitively operate the device. As such, there will be work put in place to design a functional package that can be downloaded and will run like any other application. All of this will be done with freely available open-source software in order to promote the goal of improved data management and global health. One further aim is to expand applications to other medical devices and pumps that are commercially available.

Finally, to deploy this technology in both simulated and real-world healthcare environments. An example would be to setup this device in a simulation lab which is utilized to practice critical resuscitation skills for clinicians and trainees, prior to real-world application. Once feasibility and accuracy has been assessed in the simulated environment, the system can then be deployed in a real-world critical care environment here at the Health Sciences Centre in Winnipeg or other centers. Here real-time operational limitations will be explored, and the algorithms improved as needed. All future renditions and investigations will lead to improvements in the source code, which will be made openly available as new versions arise on GitHub.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

LF was responsible for concept, design, analysis, manuscript preparation. JD was responsible for concept, design and manuscript preparation. CB, AG, and AS were responsible for manuscript composition and editing. BU was responsible for concept and manuscript editing. FZ was responsible for concept, design, analysis, manuscript preparation/editing and supervision.

FUNDING

This work was directly supported by the Manitoba Public Insurance (MPI) Neuroscience/TBI Research Endowment, the University of Manitoba Department of Surgery GFT Grant and University Research Grant Program (URGP). In addition, FAZ receives research support from the United States National Institutes of Health (NIH) through the National Institute of Neurological Disorders and Stroke (NINDS) (Grant #: R03NS114335-01), the Canadian Institutes of Health Research (CIHR) (Grant #: 432061), the Canada Foundation for Innovation (CFI) (Project #: 38583), Research Manitoba (Grant #: 3906), the University of Manitoba VPRI Research Investment Fund (RIF), the University of Manitoba Centre on Aging, and the University of Manitoba Rudy Falk Clinician-Scientist Professorship. LF is supported through the University of Manitoba—Department of Surgery GFT Research Grant,

and the University of Manitoba Office of Research Services (ORS)—University Research Grant Program (URGP). CB is supported through the Centre on Aging Fellowship at the University of Manitoba. AG is supported through the University of Manitoba Clinician Investigator Program.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fdata.2021.689358/full#supplementary-material

REFERENCES

Allan, J., Aslam, J., Belkin, N., Buckley, C., Callan, J., Croft, B., et al. (2003). Challenges in Information Retrieval and Language Modeling: Report of a Workshop Held at the Center for Intelligent Information Retrieval, University of Massachusetts Amherst, September 2002. *ACM SIGIR Forum* 37, 31. doi:10.1145/945546.945549

Aries, M. J., Czosnyka, M., Budohoski, K., Steiner, L., Lavinio, A., Kolias, A., et al. (2012). Continuous Determination of Optimal Cerebral Perfusion Pressure in Traumatic Brain Injury*. *Crit. Care Med.* 40, 2456. doi:10.1097/ccm.0b013e3182514eb6

Balestreri, M., Czosnyka, M., Steiner, L. A., Hiler, M., Schmidt, E. A., Matta, B., et al. (2005). Association between Outcome, Cerebral Pressure Reactivity and Slow ICP Waves Following Head Injury. *Acta Neurochir. Suppl.* 95, 25. doi:10.1007/3-211-32318-x_6

Barchard, K. A., and Pace, L. A. (2011). Preventing Human Error: The Impact of Data Entry Methods on Data Accuracy and Statistical Results. *Comput. Hum. Behav.* 27, 1834. doi:10.1016/j.chb.2011.04.004

Bradski, G. (2000). Home and Index OpenCV-Python Tutorials 1 Documentation. The OpenCV Library. https://opencv.org/ (Accessed August 1, 2021).

Cabella, B., Donnelly, J., Cardim, D., Liu, X., Cabeleira, M., Smielewski, P., et al. (2017). An Association between ICP-Derived Data and Outcome in TBI Patients: The Role of Sample Size. *Neurocrit. Care* 27, 103. doi:10.1007/s12028-016-0319-x

Canny, J. (1986). A Computational Approach to Edge Detection. *Ieee Trans. Pattern Anal. Mach. Intelligence*, 8 (6), 679–698. doi:10.1109/tpami.1986.4767851

Carney, N., Totten, A. M., O'Reilly, C., Ullman, J. S., Hawryluk, G. W. J., et al. (2017). Guidelines for the Management of Severe Traumatic Brain Injury. *Neurosurgery* Fourth Edition, 80, 6–15. doi:10.1227/neu.0000000000001432

Carvalho, M. C. (2021). Miau, a Microbalance Autosampler, HardwareX. 10, e00215. doi:10.1016/j.ohx.2021.e00215

Carvalho, M. C. (2013). Integration of Analytical Instruments with Computer Scripting. *J. Lab. Autom.* 18, 328. doi:10.1177/2211068213476288

Carvalho, M. C. (2016). *Optical Character Recognition Practical Laboratory Automation*. Weinheim, Germany: John Wiley & Sons, 207–209. doi:10.1002/9783527801954.app2

Chalmers, T. C., Smith, H., Jr, Blackburn, B., and Silverman, B. (1981). A Method For Assessing The Quality Of a Randomized Control Trial, Control. *Clin. Trials* 2, 31.

Cheng, H., Bouman, C. A., and Allebach, J. P. (1997). Multiscale Document Segmentation, in IS&T 50th Annual Conference, Cambridge, MA, May 18–23, 1997, pp. 417–425.

Delaney, N. F., Rojas Echenique, J. I., and Marx, C. J. (2013). Clarity: An Open-Source Manager for Laboratory Automation. *J. Lab. Autom.* 18, 171. doi:10.1177/2211068212460237

Feng, G., Bouman, C., and Cheng, H. (2006). High-Quality MRC Document Coding. *IEEE Trans. Image Process.*. doi:10.1109/tip.2006.877493

Fischer, S., Lienhart, R., and Effelsberg, W. (1995). Automatic Recognition of Film Genres." in Proceedings of the Third ACM International Conference on Multimedia. New York, NY, USA: Association for Computing Machinery, 295–304. doi:10.1145/217279.215283

Froese, L., Dian, J., Batson, C., Gomez, A., Alarifi, N., Unger, B., et al. (2020). The Impact of Vasopressor and Sedative Agents on Cerebrovascular Reactivity and Compensatory Reserve in Traumatic Brain Injury: An Exploratory Analysis. *Neurotrauma Rep.* 1, 157. doi:10.1089/neur.2020.0028

Froese, L., Dian, J., Batson, C., Gomez, A., Unger, B., and Zeiler, F. A. (2020). The Impact of Hypertonic Saline on Cerebrovascular Reactivity and Compensatory Reserve in Traumatic Brain Injury: An Exploratory Analysis. *Acta Neurochir. (Wien)* 1. doi:10.1007/s00701-020-04579-0

Hatchett, S., and Vfrolov, V. (1991). Null-Modem Emulator (Com0com) - Virtual Serial Port Driver for Windows. https://sourceforge.net/projects/com0com/ (Accessed August 1, 2021).

Jung, K., In Kim, K., and Jain, A. K. (2004). Text Information Extraction in Images and Video: A Survey. *Pattern Recognit* 37, 977. doi:10.1016/j.patcog.2003.10.012

Karthikeyan, S., Seco de Herrera, A. G., Doctor, F., and Mirza, A. (2021). An OCR Post-Correction Approach Using Deep Learning for Processing Medical Reports. *IEEE Trans. Circuits Syst. Video Technol.* 1. doi:10.1109/tcsvt.2021.3087641

Kim, S., Kim, D., Ryu, Y., and Kim, G. (2002). "A Robust License-Plate Extraction Method under Complex Image Conditions," in Proceedings of the 16 Th International Conference on Pattern Recognition (ICPR'02) Volume (USA: IEEE Computer Society), 30216.

Klein, S. P., Fieuws, S., Meyfroidt, G., and Depreitere, B. (2020). Effects of Norepinephrine, Propofol and Hemoglobin Concentration on Dynamic Measurements of Cerebrovascular Reactivity in Acute Brain Injury. *J. Neurotrauma* 38 (4), 506–512. doi:10.1089/neu.2020.7160

Le, A. D., Pham, D. V., and Nguyen, T. A. (2019). "Deep Learning Approach for Receipt Recognition," in *Future Data and Security Engineering*. Editors T. K. Dang, J. Küng, M. Takizawa, and S. H. Bui (Cham: Springer International Publishing), 705–712. doi:10.1007/978-3-030-35653-8_50

Lee, M. (2007). *Pytesseract: Python-Tesseract Is a Python Wrapper for Google's Tesseract-OCR*.

Li, H., and Doermann, D. (1999). "Text Enhancement in Digital Video Using Multiple Frame Integration," in Proceedings of the Seventh ACM International Conference on Multimedia (Part 1) - MULTIMEDIA '99 (Orlando, Florida, United States: ACM Press), 19–22. doi:10.1145/319463.319466

Li, H., Doermann, D., and Kia, O. (2000). Automatic Text Detection and Tracking in Digital Video. *IEEE Trans. Image Process.* 9, 147. doi:10.1109/83.817607

Lundh, F. (1999). An Introduction to Tkinter. Available at: Www.Pythonware.Com/Library/Tkinter/Introduction/Index.Htm.

Mardis, E. R. (2011). A Decade's Perspective on DNA Sequencing Technology. *Nature* 470, 198. doi:10.1038/nature09796

Matchett, K. B., Lynam-Lennon, N., Watson, R. W., and Brown, J. A. L. (2017). Advances in Precision Medicine: Tailoring Individualized Therapies. *Cancers (Basel)* 9. doi:10.3390/cancers9110146

Mokhtar, K., Bukhari, S. S., and Dengel, A. (2018). "OCR Error Correction: State-Of-The-Art vs an NMT-Based Approach," in 3th IAPR International Workshop on Document Analysis Systems, Vienna, Austria (DAS), 429–434. doi:10.1109/das.2018.63

Namysl, M., and Konya, I. (2019). "Efficient, Lexicon-free OCR Using Deep Learning," in International Conference on Document Analysis and Recognition, Sydney, Australia, September 20–25, 2019 (ICDAR), 295–301. doi:10.1109/icdar.2019.00055

Oakley, J. P., and Satherley, B. L. (1998). Improving Image Quality in Poor Visibility Conditions Using a Physical Model for Contrast Degradation. *IEEE Trans. Image Process.* 7, 167. doi:10.1109/83.660994

OpenCV. Canny Edge Detection. Available at: https://docs.opencv.org/trunk/da/d22/tutorial_py_canny.html.

Open. Image Thresholding Available at: https://docs.opencv.org/master/d7/d4d/tutorial_py_thresholding.html.

Park, S. H., Kim, K. I., Jung, K., and Kim, H. J. (1999). Locating Car License Plates Using Neural Networks. *Electron. Lett.* 35, 1475. doi:10.1049/el:19990977

Schantz, H. F. (1982). *The History of OCR, Optical Character Recognition*: [Manchester Center Recognition Technologies Users Association].

Steiner, L., Czosnyka, M., Piechnik, S., Smielewski, P., Chatfield, D., Menon, D., et al. (2002). Continuous Monitoring of Cerebrovascular Pressure Reactivity Allows Determination of Optimal Cerebral Perfusion Pressure in Patients with Traumatic Brain Injury. *Crit. Care Med.* 30, 733. doi:10.1097/00003246-200204000-00002

Welcome to PySerial's Documentation PySerial 3.4 Documentation. Available at: https://pyserial.readthedocs.io/en/latest/.

Wemhoener, D., Yalniz, I. Z., and Manmatha, R. (2013). "Creating an Improved Version Using Noisy OCR from Multiple Editions," in 12th International Conference on Document Analysis and Recognition, Washington, DC, August 25–28, 2013 (Washington, DC, USA: IEEE), 160–164. doi:10.1109/icdar.2013.39

Witten, I. H., Don, K. J., Dewsnip, M., and Tablan, V. (2004). Text Mining in a Digital Library. *Int. J. Digit. Libr.* 4, 56. doi:10.1007/s00799-003-0066-4

Yalniz, I. Z., and Manmatha, R. (2012). "An Efficient Framework for Searching Text in Noisy Document Images," in *10th IAPR International Workshop on Document Analysis Systems*. Gold Coast, Australia: IEEE, 48–52. doi:10.1109/das.2012.18

Yalniz, I. Z., and Manmatha, R. (2019). Dependence Models for Searching Text in Document Images. *IEEE Trans. Pattern Anal. Mach. Intell.* 41, 49. doi:10.1109/tpami.2017.2780108

Yin, Y., Zhang, W., Hong, S., Yang, J., Xiong, J., and Gui, G. (2019). Deep Learning-Aided OCR Techniques for Chinese Uppercase Characters in the Application of Internet of Things. *IEEE Access* 7, 47043. doi:10.1109/access.2019.2909401

Zeiler, F. A., Donnelly, J., Smielewski, P., Menon, D. K., Hutchinson, P. J., and Czosnyka, M. (2018). Critical Thresholds of Intracranial Pressure-Derived Continuous Cerebrovascular Reactivity Indices for Outcome Prediction in Noncraniectomized Patients with Traumatic Brain Injury. *J. Neurotrauma* 35, 1107–1115. doi:10.1089/neu.2017.5472

Zeiler, F. A., Ercole, A., Cabeleira, M., Beqiri, E., Zoerle, T., and Carbonara, M., (2020). Patient-Specific ICP Epidemiologic Thresholds in Adult Traumatic Brain Injury: A CENTER-TBI Validation Study. *J. Neurosurg. Anesthesiol* 33 (1), 28–38. doi:10.1097/ANA.0000000000000616

Zeiler, F. A, Ercole, A., Cabeleira, M., Carbonara, M., Stocchetti, N., Menon, D. K., et al. (2019). Resolution (HR ICU) Sub-study Participants and Investigators, Comparison Of Performance Of Different Optimal Cerebral Perfusion Pressure Parameters For Outcome Prediction In Adult Traumatic Brain Injury: A Collaborative European NeuroTrauma Effectiveness Research In Traumatic Brain Injury (CENTER-TBI) Study, *J. Neurotrauma* 36, 1505. doi:10.1089/neu.2018.6182

Zeiler, F. A., Lee, J. K., Smielewski, P., Czosnyka, M., and Brady, K. (2018). Validation of Intracranial Pressure-Derived Cerebrovascular Reactivity Indices against the Lower Limit of Autoregulation, Part II: Experimental Model of Arterial Hypotension. *J. Neurotrauma* 35, 2812–2819. doi:10.1089/neu.2017.5604

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership