



## OPEN ACCESS

## EDITED BY

Michał Korostynski,  
Institute of Pharmacology PAS, Poland

## REVIEWED BY

Jakub Nalepa,  
Silesian University of Technology, Poland  
Roland Goecke,  
University of Canberra, Australia

## \*CORRESPONDENCE

Stanley Sau Ching Wong,  
✉ wongstan@hku.hk

## SPECIALTY SECTION

This article was submitted to  
Computational Genomics,  
a section of the journal  
Frontiers in Genetics

RECEIVED 25 August 2022

ACCEPTED 20 March 2023

PUBLISHED 29 March 2023

## CITATION

Wang F, Cheung CW and Wong SSC  
(2023), Use of pain-related gene features  
to predict depression by support vector  
machine model in patients  
with fibromyalgia.  
*Front. Genet.* 14:1026672.  
doi: 10.3389/fgene.2023.1026672

## COPYRIGHT

© 2023 Wang, Cheung and Wong. This is  
an open-access article distributed under  
the terms of the [Creative Commons  
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,  
distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication  
in this journal is cited, in accordance with  
accepted academic practice. No use,  
distribution or reproduction is permitted  
which does not comply with these terms.

# Use of pain-related gene features to predict depression by support vector machine model in patients with fibromyalgia

Fengfeng Wang, Chi Wai Cheung and Stanley Sau Ching Wong\*

Department of Anaesthesiology, School of Clinical Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China

The prevalence rate of depression is higher in patients with fibromyalgia syndrome, but this is often unrecognized in patients with chronic pain. Given that depression is a common major barrier in the management of patients with fibromyalgia syndrome, an objective tool that reliably predicts depression in patients with fibromyalgia syndrome could significantly enhance the diagnostic accuracy. Since pain and depression can cause each other and worsen each other, we wonder if pain-related genes can be used to differentiate between those with major depression from those without. This study developed a support vector machine model combined with principal component analysis to differentiate major depression in fibromyalgia syndrome patients using a microarray dataset, including 25 fibromyalgia syndrome patients with major depression, and 36 patients without major depression. Gene co-expression analysis was used to select gene features to construct support vector machine model. The principal component analysis can help reduce the number of data dimensions without much loss of information, and identify patterns in data easily. The 61 samples available in the database were not enough for learning based methods and cannot represent every possible variation of each patient. To address this issue, we adopted Gaussian noise to generate a large amount of simulated data for training and testing of the model. The ability of support vector machine model to differentiate major depression using microarray data was measured as accuracy. Different structural co-expression patterns were identified for 114 genes involved in pain signaling pathway by two-sample KS test ( $p < 0.001$  for the maximum deviation  $D = 0.11 > D_{critical} = 0.05$ ), indicating the aberrant co-expression patterns in fibromyalgia syndrome patients. Twenty hub gene features were further selected based on co-expression analysis to construct the model. The principal component analysis reduced the dimension of the training samples from 20 to 16, since 16 components were needed to retain more than 90% of the original variance. The support vector machine model was able to differentiate between those with major depression from those without in fibromyalgia syndrome patients with an average accuracy of 93.22% based on the expression levels of the selected hub gene features. These findings would contribute key information that can be used to develop a clinical decision-making tool for the data-driven, personalized optimization of diagnosing depression in patients with fibromyalgia syndrome.

## KEYWORDS

chronic pain, gene co-expression, depression, fibromyalgia, principal component analysis, support vector machine

## 1 Introduction

Chronic pain, defined as pain lasting at least 3 months, is a major clinical and socioeconomic burden to both patients and society (Queiroz, 2013; Eucker et al., 2022). Fibromyalgia syndrome (FMS) is a chronic pain condition associated with diffuse musculoskeletal pain, cognitive dysfunction, sleep disturbance, somatic symptoms, and psychological distress (Siracusa et al., 2021; Farag et al., 2022; Kumbhare et al., 2022). FMS affects around 5% of global population, and is more prevalent in females (Wolfe et al., 1995; Siracusa et al., 2021). Patients with FMS often suffer from functional disabilities, including impairment with work and activities of daily living (Arnold et al., 2008; Clauw et al., 2011). FMS is associated with psychiatric and psychological disorders, such as depression and anxiety (Bennett et al., 2007; Clauw et al., 2011). Depression is more likely to be present in patients with FMS (Santos et al., 2011), with an incidence ranging from 9.2% to 90% depending on different screening or diagnostic methods (Wilke et al., 2010; Gota et al., 2017). It has been reported that more than half of all FMS patients experienced major depressive disorder during their life-time (Løge-Hagen et al., 2019). Patients with chronic pain who have co-existing depression suffer from greater pain intensity and longer duration of pain, and are less likely to respond to treatment (Bair et al., 2003). Depression is commonly unrecognized in patients with chronic pain, and therefore untreated (Lee et al., 2018). Negating depression in chronic pain patients may increase the chance of treatment failure. Formal diagnosis usually depends on clinical experience and subjective evaluation by psychiatrists. However, physicians managing chronic pain in FMS patients are non-psychiatrists and often lack the expertise to diagnose major depression with a high level of certainty. This leads to low consistency and accuracy, resulting in major depression in FMS patients being missed (Yan et al., 2022). A personalized tool for differentiating FMS patients with major depression by identifying biological predictors in patients with FMS will enhance diagnostic accuracy and improve clinical outcomes.

Support vector machine (SVM) is one promising technique for identifying biological markers, which has been widely used to diagnose and classify various diseases, especially depression. Researchers have developed an integrated analytical algorithm consisting of nuclear magnetic resonance-based metabolomics and least squares-SVM to diagnose depression (Zheng et al., 2017). SVM was applied to predict the efficacy of escitalopram from electroencephalography recordings for treatment of depression (Zhdanov et al., 2020). Furthermore, SVM was used to separate depressed from healthy individuals based on multiple brain network properties, such as diffusion-weighted neuroimaging and graph theory (Sacchet et al., 2015). These studies provide strong evidence that SVM can be applied to predict depression. However, they cannot translate these findings into clinical tools, due to low accuracy of prediction or small sample sizes. To overcome these limitations, the present study attempted to use Gaussian noise to generate a large amount of simulated data for model construction with a high accuracy.

Since pain and depression can cause each other and worsen each other, we hypothesize that pain-related genes can be used to differentiate between those with major depression from those without. In this study, whole genome data analysis was applied

to select gene features to construct SVM model combined with principal component analysis (PCA) technique for prediction of depression in patients with FMS. Usually, multiple variables are included in SVM models, leading to a high-dimension, and it is difficult to find the patterns in data with high-dimension. PCA can help reduce the number of data dimensions without much loss of information, and identify patterns in data easily (Jolliffe, 1986). Our proposed method could provide useful information for personalized optimization of diagnosing depression in patients with FMS.

## 2 Materials and methods

### 2.1 Microarray expression data and study subjects

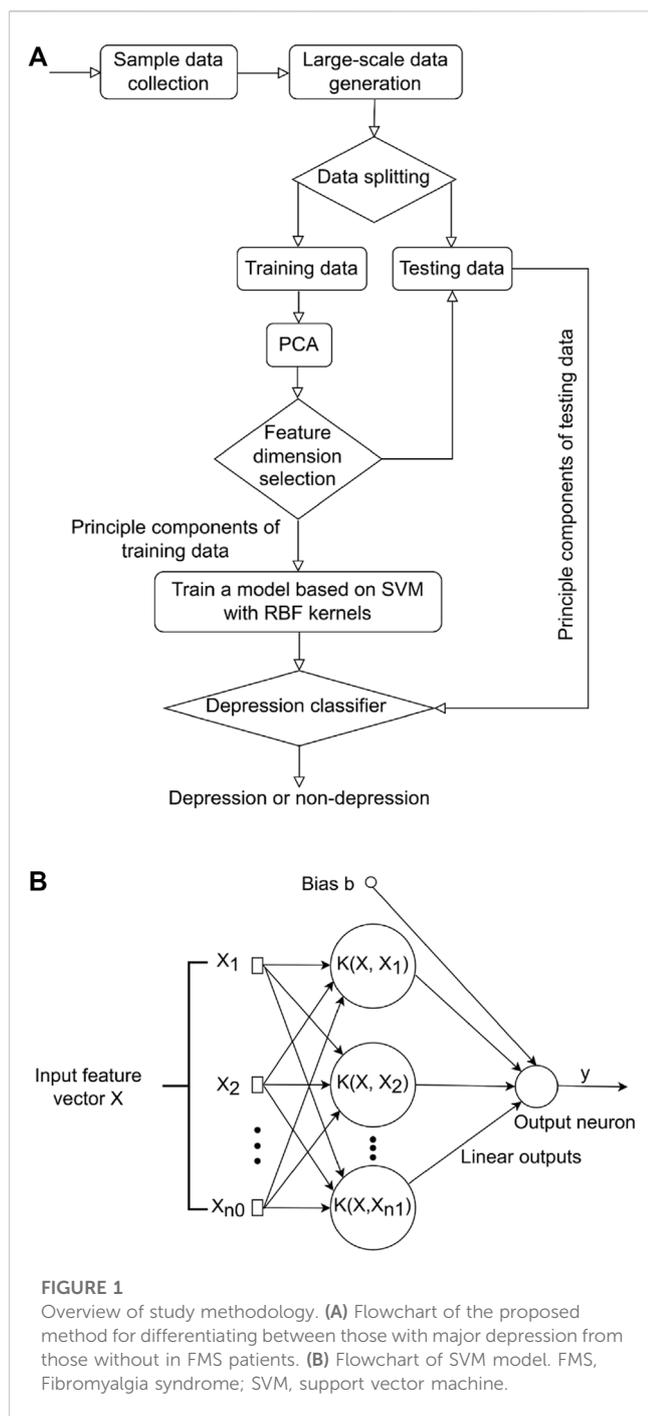
The microarray dataset GSE67311, from publicly available Gene Expression Omnibus (GEO) repository database, was used for analysis in this study (Jones et al., 2016). The data were normalized by Robust Multi-array Average method across the samples. Blood samples were collected from 70 FMS patients and 70 healthy matched controls. The subjects were limited to Caucasian females aged 18 and over. Fibromyalgia patients were diagnosed by a physician with FMS for at least 6 months or longer (Jones et al., 2016). There were 25 FMS patients with major depression, and 36 patients without major depression, which were used to construct the model.

### 2.2 Co-expression analysis and feature selection

Usually, pain and depression can cause each other and worsen each other. The pain signaling system consists of transduction, conduction, synaptic transmission, and modulation, and 114 pain-related genes identified from the microarray dataset were summarized for analysis (Supplementary Table S1) (Julius and Basbaum, 2001; Foulkes and Wood, 2008). We further extracted the expression profiles of these 114 genes from dataset GSE67311 for co-expression analysis according to the methods developed in our previous studies (Wang et al., 2014; Chan et al., 2015; Wang et al., 2015; Wang et al., 2016).

An FMS-specific cutoff point was identified to classify the co-expressed gene pairs into strong and weak co-expression classes. The normal-specific strongly co-expressed pairs were the gene pairs strongly co-expressed only in the healthy individuals, which were disrupted in the FMS group, called disrupted links. The FMS-specific strongly co-expressed pairs were the gene pairs strongly co-expressed only in the FMS group, which were invoked in the FMS group, called invoked links. There were also common weakly co-expressed pairs and common strongly co-expressed pairs. The disrupted links were regarded as the inter-gene linkages maintaining physiological balance in healthy individuals. The invoked links represented the characteristics of the disease and may be the pathogenic alternatives.

The sum of absolute Pearson correlation coefficient ( $|r|$ ) values for each gene in the co-expression network was calculated, defined as Rsum. Genes with the largest Rsum were called hub genes. Throughout the study, Rsum values were used to select hub genes for depression prediction. Cytoscape is an open source



software used to visualize complex (Shannon et al., 2003). ClueGO is a Cytoscape plug-in to annotate large sets of genes by integrating Gene Ontology, KEGG/BioCarta pathways and other databases (Bindea et al., 2009). We applied ClueGO in Cytoscape to interpret the biological function of selected hub gene features.

## 2.3 Support vector machine model construction

After a subset of gene features was selected, SVM was used to distinguish depressed and non-depressed FMS patients. PCA is a

powerful tool to reduce the number of data dimensions without much loss of information (Jolliffe, 1986). A classic method taking advantages of both SVM and PCA techniques was proposed in this study (Figure 1A). More specifically, the large-scale training and testing data sets were first available through an additive model, and then split for training and testing. PCA was subsequently adopted to reduce the feature vector dimension of the training data preceding the training process. SVM was employed to learn a base model from the low dimensional feature space to recognize depression or non-depression in FMS patients.

### 2.3.1 Generation of big-simulated gene data

The 61 samples available in the database cannot represent every possible variation of each patient. One important property of an SVM model to predict depression was the ability to generalize to the data obtained from a clinical site that was not involved in the training of the model. We adopted an additive model of embedding Gaussian noise to improve the capacity for such generalization (Formula 1). This additive model was controlled under a value of a high signal-noise-ratio (SNR) to generate a large amount of simulated data for training and testing (Wong, 2013).

$$G'_i = G_i + N_i \quad i = 1, 2, \dots \quad (1)$$

### 2.3.2 Reducing dimension of data based on PCA method

In order to reduce dimension of each sample with less loss of information, PCA technique was applied to obtain the optimal expression profiles for each sample. The sample can be expressed as a  $d$ -dimensional feature vector (Formula 2).

$$G = (g_1 \ g_2 \ g_3 \ \dots \ g_d)^T \quad (2)$$

where  $d$  denoted the feature vector dimension of one sample. For each sample, we had  $d$  observations denoting the expression profiles of multiple hub genes.

The whole processes for PCA included: i) constructing the training data set; ii) subtracting the mean from each data dimension; iii) calculating the covariance matrix; iv) calculating the eigenvectors and eigenvalues of the covariance matrix; v) determining the number of principal components; and vi) forming feature vectors for testing samples. Through PCA operation, the information redundancy of gene signals can be removed, making it easier for the model to separate depression from non-depression.

### 2.3.3 Differentiating between those with major depression from those without in patients with fibromyalgia based on SVM model

The soft-margin SVM was constructed to differentiate depressed and non-depressed patients. Let  $L$  denoted the set of sample categories, and  $L = \{1, \dots, l\}$  where  $l$  denoted the total number of sample categories ( $l = 2$ ). This problem was considered as giving  $n$  labeled empirical samples.

$$(X_1, y_1), \dots, (X_n, y_n)$$

where  $\{X_i\}_{i=1}^n$  was the PCA feature set of input samples in  $R^D$  ( $D = 20$ ), and  $y_i$  denoted the label of  $X_i$ . For the classification problem;

denoted the feature vector of one sample, an input PCA feature vector with the dimension of  $p$  ( $p = 20$ ) and  $X_i$  was then transformed to a higher dimensional feature space (Chang and Lin, 2011; Nalepa and Kawulok, 2019) through a kernel function  $K(X, X_i)$  for classification as shown (Figure 1B).

Now, the goal was to find the optimal separating hyperplane that can maximize the distance between it and the nearest data points in different categories (Formula 3).

$$f(x) = \mathbf{w} \cdot \mathbf{x} + b \quad (3)$$

The nearest points to the hyperplane were called support vectors. To find the optimal hyperplane  $f(x, w)$ , only support vectors were considered in SVM. For two linearly separated classes, the training data must satisfy the following two constraints:

$$\begin{aligned} \mathbf{w}^T \mathbf{x}_i + b &\geq 1 & \text{if } y_i = 1 \\ \mathbf{w}^T \mathbf{x}_i + b &\leq -1 & \text{if } y_i = -1 \end{aligned} \quad (4)$$

The margin of separation between two classes was thereby calculated as follows:

$$\rho = \frac{2}{\|\mathbf{w}\|}$$

To separate depression from non-depression, the problem was to maximize the margin  $\rho$  under the constraints (Formula 4). This can be obtained by solving the optimization problem during the training process (Cortes and Vapnik, 1995). By introducing Lagrange multipliers  $\{\alpha_i\}_{i=1}^N$  and dual transformation (Scholkopf et al., 1997; Duan and Keerthi, 2005; Chang and Lin, 2011), the category of hub genes in the test set was predicted as follows (Formula 5):

$$\begin{aligned} d_i &= \text{sign}(\mathbf{w}^T \mathbf{x}_i) = \text{sign}\left(\sum_{j \in SV} \alpha_j d_j \mathbf{x}_j \cdot \mathbf{x}_i\right) \\ &= \text{sign}\left(\sum_{j \in SV} \alpha_j d_j k(\mathbf{x}_j, \mathbf{x}_i)\right) \end{aligned} \quad (5)$$

where  $SV$  denoted the set of support vectors,  $k(\mathbf{x}_j, \mathbf{x}_i)$  was a kernel function, and it was the inner product of two feature vectors. By using this kernel function, the training samples can be mapped from an input space to another feature space, which increased separation of the samples. There are two commonly used kernels for SVM in real applications, namely, linear, and radial basis function (RBF) (Apostolidis-Afentoulis and Lioufi, 2015). Generally, in the case of linearly separable data, linear kernels and RBF kernels can show similar performance, while in the case of linearly non-separable data, RBF kernels can give better prediction. Note that the dataset GSE67311 was linear non-separable data, and the number ( $p$ ) of feature vector was also smaller than the number ( $n$ ) of training data. For the linearly non-separable data, SVM with RBF kernels had better performance in our study. The experimental results also illustrated that the model with RBF kernels presented the better accuracy than that with linear kernels.

To obtain a more robust model with RBF kernels for the prediction of depression, K-fold cross-validation and grid search were implemented to obtain the optimal values of  $\gamma$  and the penalty parameter ( $C$ ). In the implementation, the precision was observed for 6 splits and 6 repeats. For each iteration, the training dataset ( $trainSetAll$ ) was split into 6 folds, where 5 sets were served as the  $trainSet$ , and the remaining 1-fold was served as the  $testSet$ . The prediction accuracy of the model was calculated on the  $testSet$  in

each iteration. Finally,  $\gamma$  and  $C$  were obtained with the highest cross-validation accuracy. The best  $\gamma$  and  $C$  were then used to train the whole training set, and the final model was generated to predict the whole testing set.

**Input:**  $trainSetAll$ ,  $type$ ,  $kernel$ ,  $m1$ ,  $m2$

**Output:**  $bestC$  and  $bestGam$

**begin**

$(trainSet, testSet) \leftarrow KfoldSplit(trainSetAll)$ ;

$bestAccuracy = 0$ ;

$bestC = 0$ ;

$bestGam = 0$ ;

**for**  $c = 2^{(-m1)}:2^{(m1)}$

**for**  $s = 2^{(-m2)}:2^{(m2)}$

$model \leftarrow SVMtrain(kernel, trainSet, bestC, bestGam)$ ;

$acc \leftarrow SVMtest(model, testSet)$ ;

**if** ( $acc > bestAccuracy$ )

$bestAccuracy = acc$ ;  $bestC = c$ ;  $bestGam = s$ ;

**end**

**end**

**end**

**end**

## 2.4 Statistical analysis

The co-expression analysis and SVM model construction were conducted using MATLAB R2022a. The statistical analysis was performed by IBM SPSS Statistics 27.0 software. Chi-square test was used to determine if genes were more disrupted or invoked in FMS patients. The normality of the data was checked first. The independent sample  $t$ -test was applied to exam the statistical significance for  $R_{sum}$  values. Significant differences were found at  $p$ -value  $< 0.05$ .

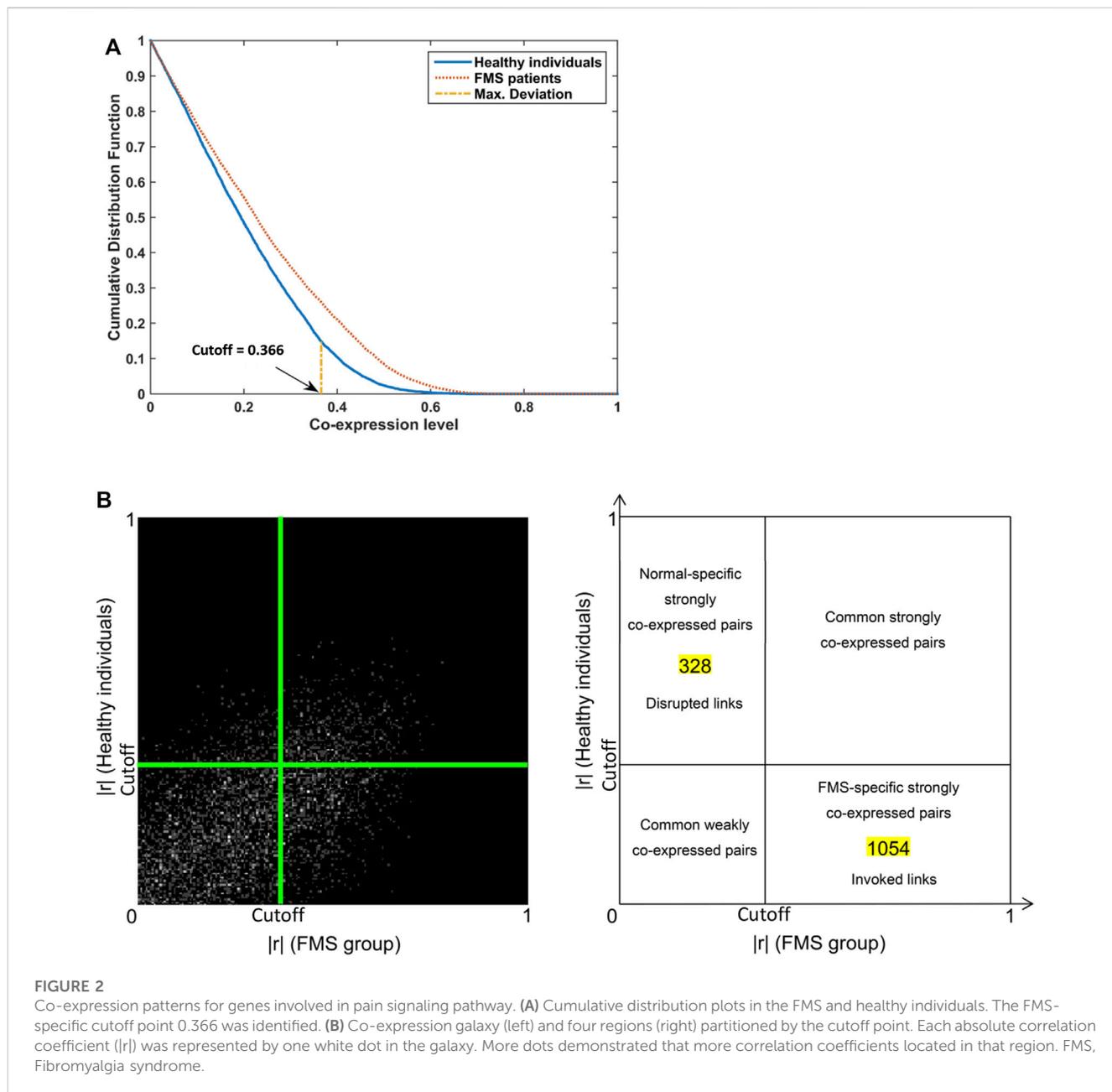
## 3 Results

### 3.1 Structural Co-expression pattern and galaxy

We calculated the correlation coefficients of 114 pain pathway genes in the FMS and healthy individuals, respectively. Two-sample KS test identified the significant difference in these two cumulative distributions with  $p$ -value  $< 0.001$  for the maximum deviation  $D = 0.11 > D_{critical} = 0.05$ . The FMS-specific cutoff point, 0.366, was identified at the maximum deviation (Figure 2A). The co-expression galaxy was plotted and partitioned into four regions: i) Disrupted links; ii) Common strongly co-expressed pairs; iii) Invoked links; and iv) Common weakly co-expressed pairs (Figure 2B). From the results, we observed that links were more likely to be invoked than disrupted in FMS group, Chi-square test,  $p < 0.001$ .

### 3.2 Identification and annotation of hub gene features

There were 328 disrupted and 1,054 invoked links, of which involved genes were strongly co-expressed in the healthy individuals



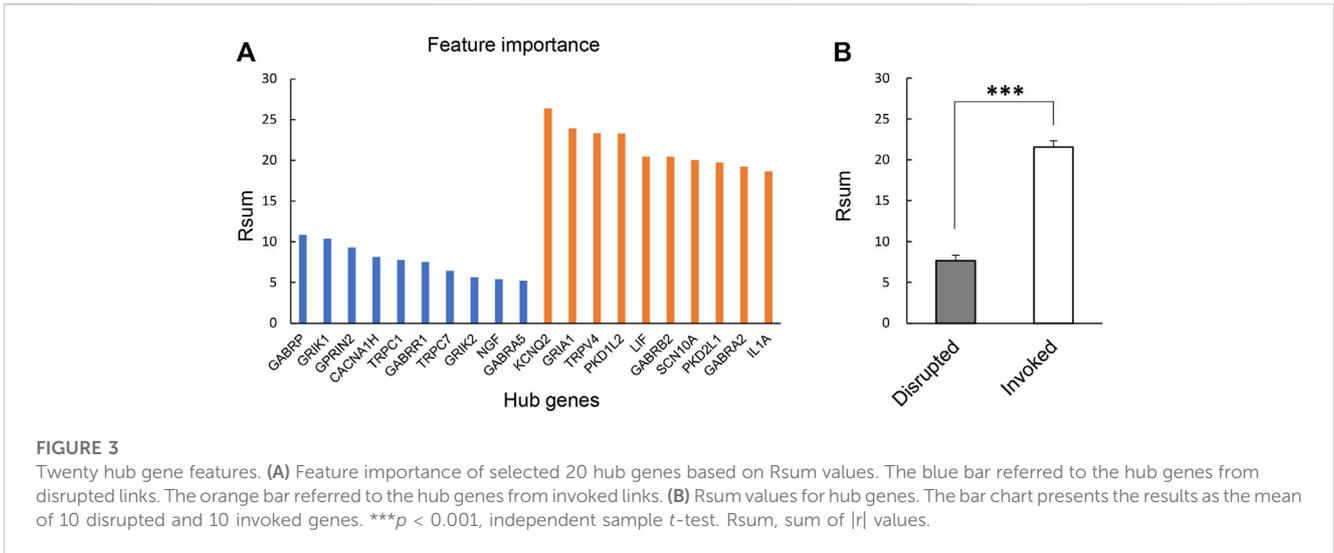
or FMS group (Figure 2B). We further explored the hub genes with the highest degree of connectivity in each group. The most important features contributing to the model were selected based on the Rsum values, including top 10 invoked and top 10 disrupted hub genes (Figure 3A). The Rsum values were significantly larger for invoked links than disrupted links ( $t$ -test,  $p < 0.001$ ), indicating the important role of invoked hub genes in the development of FMS (Figure 3B).

Four major significant biological function groups were found to annotate the selected hub gene features by ClueGO in Cytoscape (Table 1; Supplementary Figure S1A): i) Ion channel complex, adjusted  $p$ -value  $< 0.001$ , 13 involved genes; ii) TRPs channels, adjusted  $p$ -value  $< 0.001$ , 3 involved genes; iii) Sodium channel activity, adjusted  $p$ -value  $< 0.001$ , 5 involved genes; and iv) Transmitter-gated ion channel activity involved in regulation of

postsynaptic membrane potential, adjusted  $p$ -value  $< 0.001$ , 9 involved genes. Each major group was composed of one or several terms/pathways. The group of transmitter-gated ion channel activity involved in regulation of postsynaptic membrane potential had the most terms/pathways, and the percentage was the highest, 86.05% (Supplementary Figure S1B). The results demonstrated that these pain-related hub genes had important biological functions.

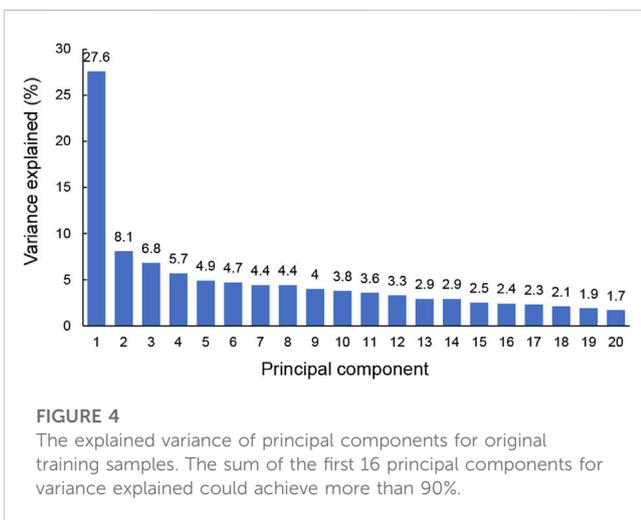
### 3.3 Differentiating between those with major depression from those without using SVM model

In this simulation, PCA was applied to 20 hub genes with expression measurements for original training samples. To get a

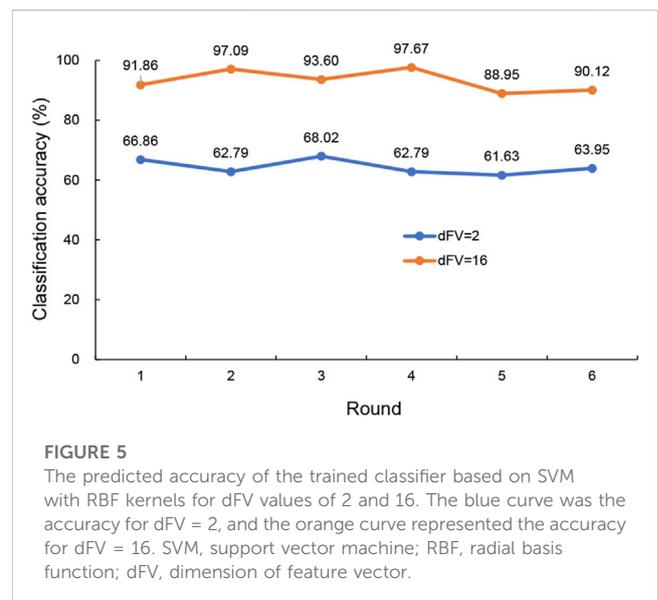


**TABLE 1 Functional annotation for the selected 20 hub genes.**

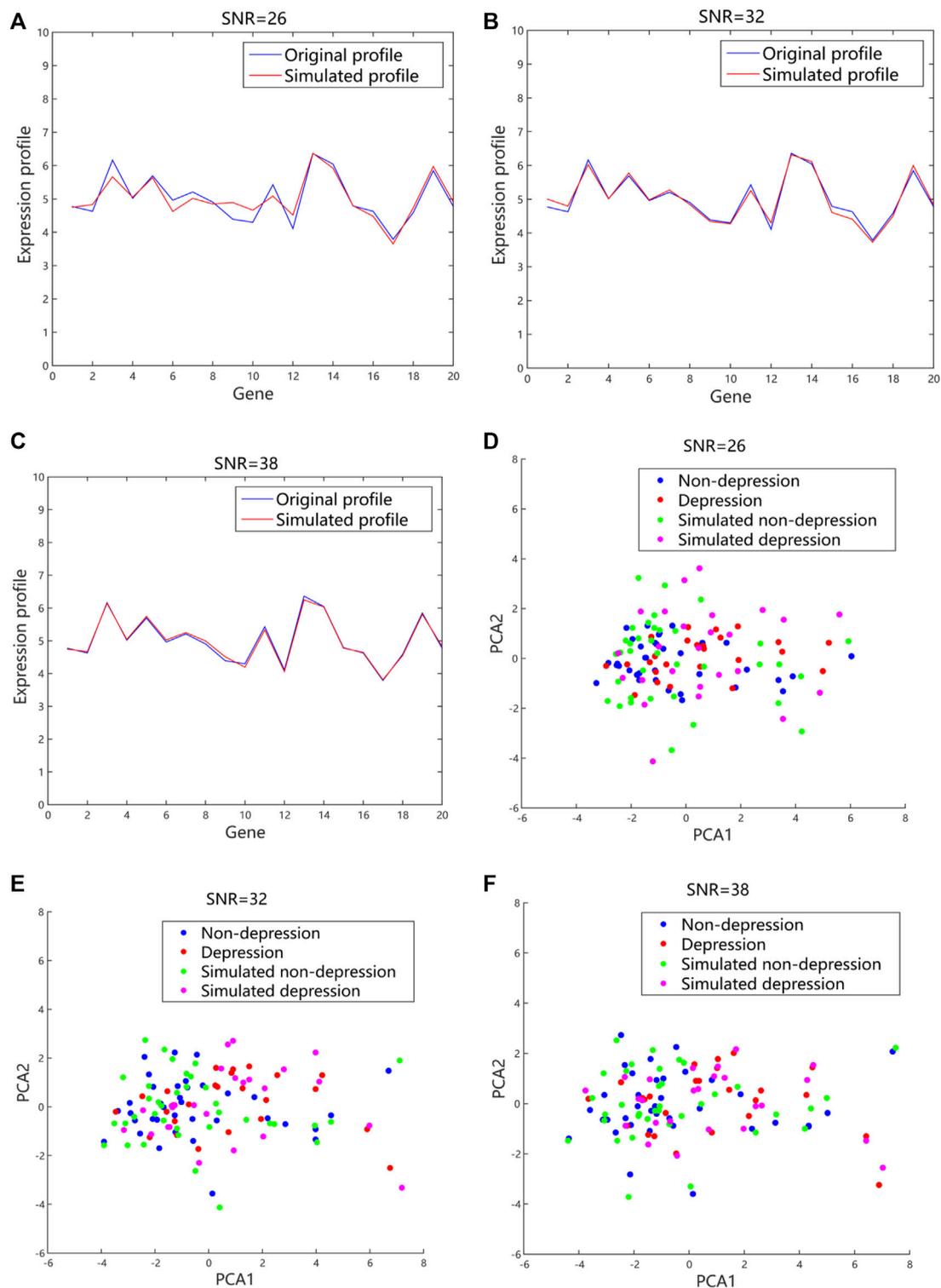
Biological function	Adjusted $p$ -value	Involved genes
Ion channel complex	2.87E-25	CACNA1H, GABRA2, GABRA5, GABRB2, GABRP, GABRR1, GRIA1, GRIK2, KCNQ2, PKD2L1, SCN10A, TRPC1, TRPC7
TRPs channels	9.46E-8	TRPC1, TRPC7, TRPV4
Sodium channel activity	5.44E-10	CACNA1H, GRIK1, GRIK2, PKD2L1, SCN10A
Transmitter-gated ion channel activity involved in regulation of postsynaptic membrane potential	3.06E-12	GABRA2, GABRA5, GABRB2, GABRP, GABRR1, GRIA1, GRIK1, GRIK2, NGF



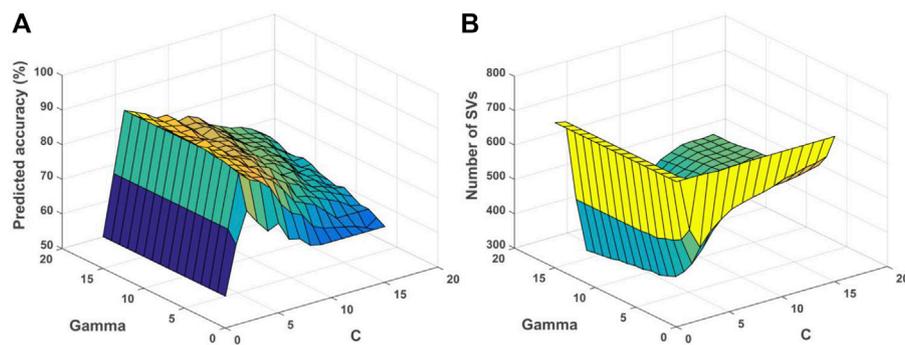
view for the dimension of the data, the proportion of the variance within each principal component was shown (Figure 4). The results demonstrated that 16 components were needed to retain more than 90% of the original variance. In such a way, PCA reduced the dimension of the training samples from 20 to 16. The similarities and differences of the feature vectors in terms of the PCA



representation could be easily highlighted due to the orthogonality properties of PCA eigenvectors. In this simulation, the trained model with the dimension of feature vector (dFV) of 16 was chosen automatically (Figure 4). We also plotted the

**FIGURE 6**

The generated gene expression data and distributions in 2D PCA space with different values of SNR based on the additive model. (A), (B), and (C) The representative simulated expression profiles with SNR values of 26, 32 and 38, respectively. The blue lines referred to the original expression profiles from microarray datasets. The red lines were the simulated expression profiles. (D), (E), and (F) The representative distributions with SNR values of 26, 32 and 38, respectively. The blue and red dots represented the original data from microarray datasets. The green and purple dots were the simulated data. 2D, 2-dimension; PCA, principal component analysis; SNR, signal-noise-ratio.



**FIGURE 7**

Relationship among parameters, predicted accuracy and support vectors for the depression prediction. **(A)** Relationship among *gamma*, *C* and predicted accuracy. **(B)** Relationship among *gamma*, *C* and number of SVs. SVs, support vectors.

prediction accuracy when dFV was set to 2, and the accuracy was very low due to more information lost. By contrast, the best model with dFV of 16 achieved superior performance with the average accuracy of 93.22% to differentiate between those with major depression from those without (Figure 5).

The simulated expression profiles and their distributions in 2D PCA space with different values of SNR were shown (Figure 6). The fluctuation of gene expression profile was quite large with a small value of SNR when compared to a large value. The data were dispersed randomly due to the large noises introduced. By contrast, the generated expression profiles had small fluctuation and regular distribution when a higher value of SNR was used. Usually, the larger the SNR is, the smaller the noise mixed in the signal is. However, if SNR is too large, the generated data are quite close to the original data, and difficult to simulate more possibilities. According to our simulated data, the quality of the generated data was better with SNR value of 30 dB–40 dB compared to the value of SNR between 20 dB and 30 dB. To strike the balance between the fidelity and generalizability, the SNR was set to 32 in our proposed method.

The relationship among *gamma*, *C* and accuracy was visualized (Figure 7A). It can be observed that the value of *gamma* was larger, and the training accuracy can be higher. However, it would influence the generalization performance of each RBF kernel with the corresponding support vectors (SVs), and lead to overfitting for the training data. For unknown samples, the classification performance would be poor. By contrast, if the value of *gamma* was smaller, the smoothing effect of RBF was larger. It was difficult to obtain higher classification accuracy as shown (Figure 7A). Note that the penalty parameter *C* denoted the error intolerable of the model. The smaller the *C* was, the less fitting it was. By contrast, if *C* was too large, it was easy to make the model overfitting. Therefore, the generalization performance was poor if *C* was too large or too small as shown (Figure 7A). The relationship among *gamma*, *C* and the number of SVs was shown (Figure 7B). The smaller the *gamma* was, the number of SVs was larger, while the number of SVs would influence the time of SVM model training and testing.

The performance for SVM with RBF kernels and linear kernels was evaluated. Random Field based method was also applied for the

prediction of depression in patients with Fibromyalgia (Criminisi et al., 2011; Karpathy, 2011). The results illustrated that the model with RBF kernels presented the optimal performance among the three algorithms (Figure 8A). For the SVM model with RBF kernels, two multifold cross-validation schemes were designed. In Scheme 1, all samples including original and synthesized data were split randomly for training and testing. In Scheme 2, synthesized data were used for training the SVM model, and original data were adopted for testing. The prediction accuracy was presented (Figure 8B). From the results, we can see that the trained model had a robust and consistent performance. The optimal values of *gamma* and *C* for the 6-fold cross-validation were obtained (Figure 8C). The consuming time during the training and testing processes for the 6-fold cross-validation was shown (Figure 8D). The simulation results demonstrated that the prediction was very fast, and the 172 samples only took 0.002 s for prediction.

## 4 Discussion

In this study, we investigated the feasibility of differentiating FMS patients with major depression from those without depression using microarray data by applying the support vector machine model. Of importance, the differentiation was based on gene expression levels. This is distinct from conventional clinical practice, where major depression is identified clinically without the aid of objective biological markers.

The functional annotation showed that the pain-related hub genes had important biological functions. Thirteen hub genes were involved in ion channel complex: CACNA1H, GABRA2, GABRA5, GABRB2, GABRP, GABRR1, GRIA1, GRIK2, KCNQ2, PKD2L1, SCN10A, TRPC1, and TRPC7 (Table 1). Dysfunction in the excitation/inhibition balance could lead to the major depression (Fee et al., 2017). Ion channels play a vital role in regulating the excitability, network activity and plasticity, which can alter the GABAergic and glutamatergic neuron excitability and firing, to change the excitation/inhibition balance in microcircuits (Eren-Koçak and Dalkara, 2021). Drugs targeting ion channels, including voltage-gated Na<sup>+</sup> (VGSCs) and Ca<sup>2+</sup> channels

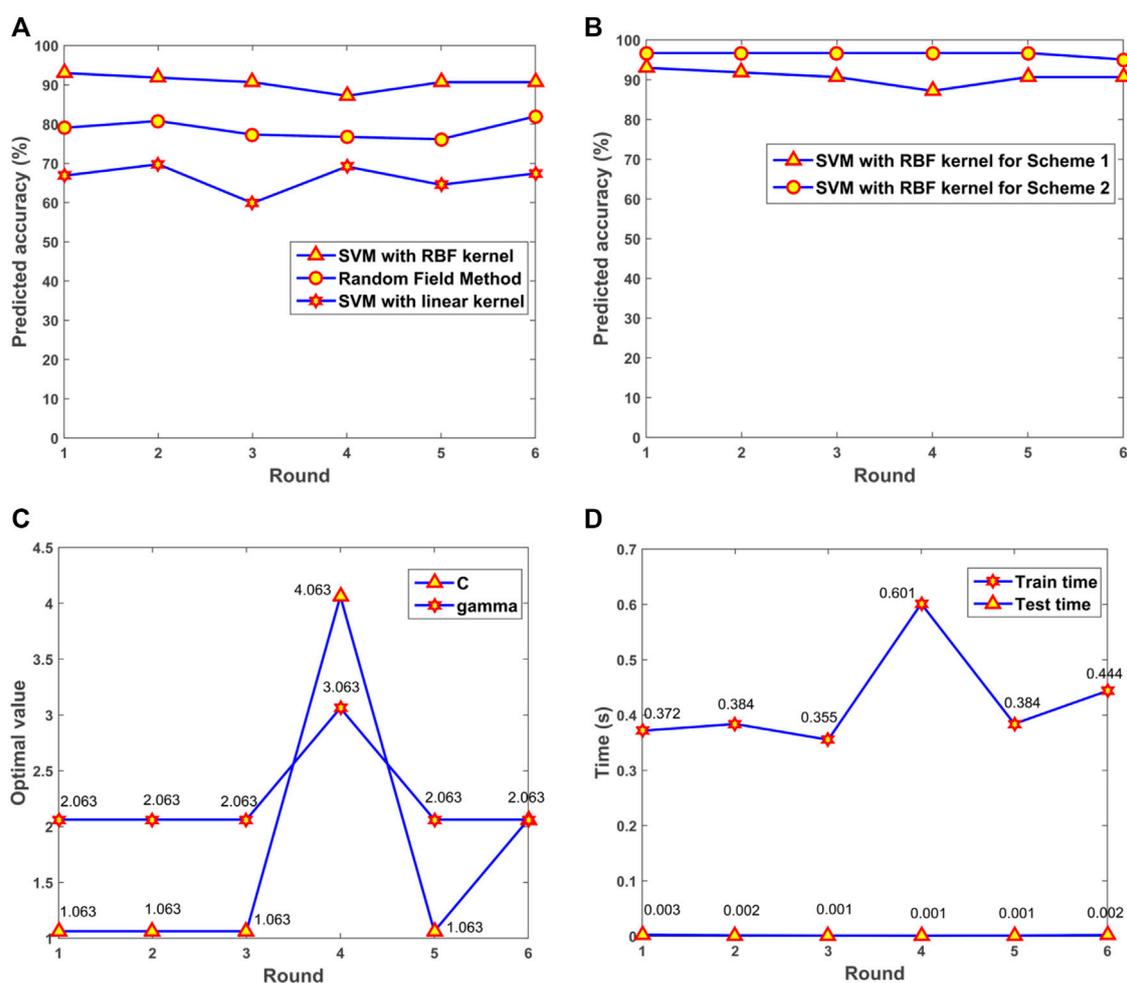


FIGURE 8

The performance of the model. (A) Predicted accuracy of the model with different kernels. (B) Predicted accuracy with different cross-validation schemes. (C) The optimal  $\gamma$  and C in the 6-fold cross-validation. (D) Consuming time for training and testing processes. SVM, support vector machine; RBF, radial basis function.

(VGCCs), are the major treatment for chronic neuropathic pain (Markman and Dworkin, 2006). Hub genes of TRPC1, TRPC7 and TRPV4 were found to be involved in TRPs transport extracellular  $\text{Ca}^{2+}$  to cytosol process (Table 1). Transient receptor potential (TRP) is a kind of cation channels expressed in non-excitatory and excitable cells (Duitama et al., 2020). TRPs located in plasma membrane could help  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Ca}^{2+}$ , and  $\text{Mg}^{2+}$  ions, and trace metal ions into the cells (Nilius and Owsianik, 2011). Some subfamily members that highly expressed in neurons and microglia could mediate the neuropathic pain (Haraguchi et al., 2012). As TRP channels are related to intracellular calcium regulation, signaling and painful stimuli transduction, they are regarded as promising targets to treat neurodegenerative diseases and pain (Duitama et al., 2020). TRP channels play a vital role in nociceptive, neuropathic, and inflammatory pain, owing to various family members participated in pain pathways (Hung and Tan, 2018). Another five hub genes, including CACNA1H, GRIK1, GRIK2, PKD2L1, and SCN10A, were associated with sodium channel activity (Table 1). Sodium channel

is one of Voltage-gated ion channels that are very important in the electrical signaling of cells (Kandel et al., 2000; de Lera Ruiz and Kraus, 2015). Genetic and functional studies found that peripheral sensory neurons expressed sodium channels were related to human pain disorders, which can be regarded as targets for the development of new analgesics (Dib-Hajj et al., 2017). Nine genes of GABRA2, GABRA5, GABRB2, GABRP, GABRR1, GRIA1, GRIK1, GRIK2, and NGF were related to Transmitter-gated ion channel activity involved in regulation of postsynaptic membrane potential (Table 1). The transmitter-gated ion channels consist of multiple subunits of membrane-spanning receptors responsible for rapid signal transduction (Barnard, 1992). These important hub genes were selected for SVM model construction to predict depression in FMS patients.

Depression is a serious mental disorder characterized by severe and persistent low mood, negative cognitive effects and behavioral symptoms (Costi et al., 2021). FMS and depression have a bidirectional relationship, where worsening of one condition

exacerbates the other, and *vice versa* (Chang et al., 2015). A number of the selected hub genes in our study have been previously reported to be related to FMS or depression, including GRIK1&2, NGF, KCNQ2, GRIA1, TRPV4, IL1A, and GABAA receptors. The hub gene was defined as the gene with the highest degree of connectivity in the co-expression network, which may play important roles in the development and characteristics of diseases, e.g., pain and depression. The expression levels of GRIK1 and GRIK2 were reported to be higher in the female patients with major depressive disorder (Gray et al., 2015). The upregulation of brain-derived neurotrophic factor mediated by NGF participates in the pathophysiological processes, leading to long-term neuroplastic changes in persistent chronic pains, e.g., fibromyalgia (Sarchielli et al., 2007). The KCNQ2/3 potassium channel is regarded as a novel treatment target for depression from preclinical studies (Costi et al., 2021). It was reported that glutamate receptors, e.g., GRIA1, may be deregulated in fibromyalgia patients (Garcia-Quintanilla and Miranzo-Navarro, 2016). TRPV4 is a kind of calcium-permeable non-selective cation channel mediating various disease states, and animals lacking TRPV4 had the decreased depression-like behavior (White et al., 2016). It was reported that significantly decreased IL1A expression was identified in depression cases compared to controls (Morrison et al., 2019). The increased GABAA receptor concentration was found in FMS compared with controls, which may lead to pain symptoms and imbalance between neuronal excitation and inhibition in FMS (Pomares et al., 2020). Our study also identified additional hub genes that have not been previously reported, such as CACNA1H and SCN10A. This discovery of hub genes that could differentiate between those with major depression from those without was based on the method that utilized both SVM and PCA techniques with a relatively high average accuracy.

The SVM studies using microarray data to select features for differentiating between those with major depression from those without are scarce. An SVM algorithm based on nuclear magnetic resonance metabolomics was developed to diagnose depression (Zheng et al., 2017). Electroencephalography recordings were applied to construct SVM classifier to predict escitalopram treatment outcome in depressed patients with the accuracy from 79.2% to 82.4% (Zhdanov et al., 2020). Moreover, diffusion-weighted neuroimaging and graph theory were used for SVM model construction to separate depressed from healthy individuals with 71.88% general accuracy (Sacchet et al., 2015). Compared to the above-mentioned methods, our current approach has the following potential advantages: i) based on the expression profiles of selected hub genes that are related to FMS or depression, making the features in SVM model more relevant to depression diagnosis in FMS patients; and ii) combination of SVM and PCA techniques to construct the model, achieving a higher accuracy of 93.22% to differentiate between those with major depression from those without.

This study has several limitations. One of these was the relatively small sample size. However, a number of other published studies using SVM for clinical medicine had even smaller sample sizes, e.g., 55 participants (Wu et al., 2018),

and 32 participants (Sacchet et al., 2015). In order to address this problem, we added Gaussian noises in each gene profile of one patient by controlling SNR (Wong, 2013) to obtain some simulated data in order to represent every possible variation of each patient. A second limitation was that although we applied various measures to control the overfitting during machine learning, such as Gaussian noise and hub gene features selection techniques, our model was not perfectly unbiased. A third possible limitation was that although this study generated a large amount of simulated data for training and testing, only one data set was used, and may need a strictly independent data set to validate the model. In a study using SVM classifier to predict escitalopram treatment outcome in depressed patients, also only one dataset was applied to construct the model (Zhdanov et al., 2020). Compared to their method, our model can achieve a higher accuracy to predict depression. To polish the predictive performance of our model, future studies should validate it with an independent data set that was not used in the process of model construction. Moreover, all the depressed patients belong to major depression in the dataset used in our study, however, it was not explicitly stated on the detailed diagnostic methods in the source article. Although this was most probably corresponded to major depressive disorder in the DSM V criteria. In addition, the prediction model demonstrated in this study has yet to be proven to be generalizable to other control groups, including depressed patients without FMS or with other kinds of pain. Model evaluation should be conducted independently with some other pain conditions to exam the generalizability. This evaluation may also help to identify whether the selected hub gene features in this study were specific to depressed FMS patients or whether they can be applied to other pain conditions.

The current model provides a base which can be further polished and improved by inputting additional information from patient blood samples. A computer-based interface system using single drop blood samples can be developed to help differentiate between those with major depression from those without in FMS patients based on the expression levels of the 20 hub genes. The aim of the proposed method was to help non-psychiatrists identify FMS patients with depression that is potentially clinically significant (e.g., those with major depression). These patients should then be referred to a clinical psychiatrist for formal diagnosis (e.g., different subtypes of depressive disorders) and subsequent management for their psychiatric condition. Given that depression is a common major barrier in the management of patients with FMS, and is also often clinically unrecognized, such computer-based interface systems could become very useful in aiding clinical diagnosis and management.

In conclusion, we have demonstrated a proof-of-concept pipeline for differentiating between those with major depression from those without among FMS patients. When developed into a proper clinical application, it may contribute to the diagnosis and clinical management for patients with co-existing FMS and depression. These findings would help to develop a clinical decision-making tool for data-driven, personalized optimization of diagnosing depression in patients with FMS.

## Data availability statement

Publicly available data were analyzed in this study. This dataset can be found here: Microarray dataset GSE67311, from publicly available Gene Expression Omnibus (GEO) repository database.

## Author contributions

FW and SW initiated the project and participated in its design. FW designed the algorithm and performed the data analyses. CC and SW participated in the data interpretation and coordination of the study. FW was responsible for writing the manuscript. All the authors participated in the discussion and editing of the manuscript.

## Funding

This work was supported by Department of Anaesthesiology, School of Clinical Medicine, Li Ka Shing Faculty of Medicine, The University of Hong Kong.

## References

- Apostolidis-Afentoulis, V., and Lioufi, K.-I. (2015). SVM classification with linear and RBF kernels. July: 0-7. Available at: [https://www.academia.edu/13811676/SVM\\_Classification\\_with\\_Linear\\_and\\_RBF\\_kernels](https://www.academia.edu/13811676/SVM_Classification_with_Linear_and_RBF_kernels).
- Arnold, L. M., Crofford, L. J., Mease, P. J., Burgess, S. M., Palmer, S. C., Abetz, L., et al. (2008). Patient perspectives on the impact of fibromyalgia. *Patient Educ. Couns.* 73 (1), 114–120. doi:10.1016/j.pec.2008.06.005
- Bair, M. J., Robinson, R. L., Katon, W., and Kroenke, K. (2003). Depression and pain comorbidity: A literature review. *Archives Intern. Med.* 163 (20), 2433–2445. doi:10.1001/archinte.163.20.2433
- Barnard, E. A. (1992). Receptor classes and the transmitter-gated ion channels. *Trends Biochem. Sci.* 17 (10), 368–374. doi:10.1016/0968-0004(92)90002-q
- Bennett, R. M., Jones, J., Turk, D. C., Russell, I. J., and Matallana, L. (2007). An internet survey of 2,596 people with fibromyalgia. *BMC Musculoskelet. Disord.* 8 (1), 27–11. doi:10.1186/1471-2474-8-27
- Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., et al. (2009). ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 25 (8), 1091–1093. doi:10.1093/bioinformatics/btp101
- Chan, L. W., Lin, X., Yung, G., Lui, T., Chiu, Y. M., Wang, F., et al. (2015). Novel structural co-expression analysis linking the NPM1-associated ribosomal biogenesis network to chronic myelogenous leukemia. *Sci. Rep.* 5 (1), 10973. doi:10.1038/srep10973
- Chang, C.-C., and Lin, C.-J. (2011). Libsvm: A library for support vector machines. *ACM Trans. intelligent Syst. Technol. (TIST)* 2 (3), 1–27. doi:10.1145/1961189.1961199
- Chang, M.-H., Hsu, J.-W., Huang, K.-L., Su, T.-P., Bai, Y.-M., Li, C.-T., et al. (2015). Bidirectional association between depression and fibromyalgia syndrome: A nationwide longitudinal study. *J. Pain* 16 (9), 895–902. doi:10.1016/j.jpain.2015.06.004
- Clauw, D. J., Arnold, L. M., and McCarberg, B. H. (2011). “The science of fibromyalgia,” in *Mayo clinic proceedings* (Amsterdam: Elsevier), 907–911.
- Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* 20 (3), 273–297. doi:10.1007/bf00994018
- Costi, S., Morris, L. S., Kirkwood, K. A., Hoch, M., Corniquel, M., Vo-Le, B., et al. (2021). Impact of the KCNQ2/3 channel opener ezogabine on reward circuit activity and clinical symptoms in depression: Results from a randomized controlled trial. *Am. J. Psychiatry* 178 (5), 437–446. doi:10.1176/appi.ajp.2020.20050653
- Criminisi, A., Shotton, J., and Konukoglu, E. (2011). *Decision forests for classification, regression, density estimation, manifold learning and semi-supervised learning*. Netherlands: Now Publishers. [internet]. Microsoft Research.
- de Lera Ruiz, M., and Kraus, R. L. (2015). Voltage-gated sodium channels: Structure, function, pharmacology, and clinical indications. *J. Med. Chem.* 58 (18), 7093–7118. doi:10.1021/jm501981g

## 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.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2023.1026672/full#supplementary-material>

- Dib-Hajj, S. D., Geha, P., and Waxman, S. G. (2017). Sodium channels in pain disorders: Pathophysiology and prospects for treatment. *Pain* 158 (1), S97–S107. doi:10.1097/j.pain.0000000000000854
- Duan, K.-B., and Keerthi, S. S. (2005). “Which is the best multiclass SVM method? An empirical study,” in *Multiple classifier systems* (Berlin: Springer), 278–285.
- Duitama, M., Vargas-López, V., Casas, Z., Albarracín, S. L., Sutachan, J.-J., and Torres, Y. P. (2020). TRP channels role in pain associated with neurodegenerative diseases. *Front. Neurosci.* 14, 782. doi:10.3389/fnins.2020.00782
- Eren-Koçak, E., and Dalkara, T. (2021). Ion Channel dysfunction and neuroinflammation in migraine and depression. *Front. Pharmacol.* 12, 777607. doi:10.3389/fphar.2021.777607
- Eucker, S. A., Knisely, M. R., and Simon, C. (2022). Nonopioid treatments for chronic pain—integrating multimodal biopsychosocial approaches to pain management. *JAMA Netw. Open* 5 (6), e2216482. doi:10.1001/jamanetworkopen.2022.16482
- Farag, H. M., Yunusa, I., Goswami, H., Sultan, I., Doucette, J. A., and Eguale, T. (2022). Comparison of amitriptyline and us food and drug administration-approved treatments for fibromyalgia: A systematic review and network meta-analysis. *JAMA Netw. Open* 5 (5), e2212939. doi:10.1001/jamanetworkopen.2022.12939
- Fee, C., Banasr, M., and Sibille, E. (2017). Somatostatin-positive gamma-aminobutyric acid interneuron deficits in depression: Cortical microcircuit and therapeutic perspectives. *Biol. psychiatry* 82 (8), 549–559. doi:10.1016/j.biopsych.2017.05.024
- Foulkes, T., and Wood, J. N. (2008). Pain genes. *PLoS Genet.* 4 (7), e1000086. doi:10.1371/journal.pgen.1000086
- Garcia-Quintanilla, A., and Miranzo-Navarro, D. (2016). Extraintestinal manifestations of celiac disease: 33-mer gliadin binding to glutamate receptor GRIN1A as a new explanation. *BioEssays* 38 (5), 427–439. doi:10.1002/bies.201500143
- Gota, C. E., Kaouk, S., and Wilke, W. S. (2017). The impact of depressive and bipolar symptoms on socioeconomic status, core symptoms, function and severity of fibromyalgia. *Int. J. Rheumatic Dis.* 20 (3), 326–339. doi:10.1111/1756-185X.12603
- Gray, A., Hyde, T., Deep-Soboslay, A., Kleinman, J., and Sodhi, M. (2015). Sex differences in glutamate receptor gene expression in major depression and suicide. *Mol. psychiatry* 20 (9), 1139–1068. doi:10.1038/mp.2015.114
- Haraguchi, K., Kawamoto, A., Isami, K., Maeda, S., Kusano, A., Asakura, K., et al. (2012). TRPM2 contributes to inflammatory and neuropathic pain through the aggravation of pronociceptive inflammatory responses in mice. *J. Neurosci.* 32 (11), 3931–3941. doi:10.1523/JNEUROSCI.4703-11.2012
- Hung, C.-Y., and Tan, C.-H. (2018). “TRP channels in nociception and pathological pain,” in *Advances in pain research: Mechanisms and modulation of chronic pain* (Springer), 13–27.

- Jolliffe, I. (1986). "Generalizations and adaptations of principal component analysis," in *Principal component analysis* (Springer), 223–234.
- Jones, K. D., Gelbart, T., Whisenant, T. C., Waalen, J., Mondala, T. S., Iklé, D. N., et al. (2016). Genome-wide expression profiling in the peripheral blood of patients with fibromyalgia. *Clin. Exp. Rheumatology* 34 (96), 89–S98.
- Julius, D., and Basbaum, A. I. (2001). Molecular mechanisms of nociception. *Nature* 413 (6852), 203–210. doi:10.1038/35093019
- Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S., Hudspeth, A. J., and Mack, S. (2000). *Principles of neural science*. New York: McGraw-Hill.
- Karpathy, A. (2011). Random-forest-matlab. Available at: <https://github.com/karpathy/Random-Forest-Matlab>.
- Kumbhare, D., Hassan, S., Diep, D., Duarte, F. C., Hung, J., Damodara, S., et al. (2022). Potential role of blood biomarkers in patients with fibromyalgia: A systematic review with meta-analysis. *Pain* 163 (7), 1232–1253. doi:10.1097/j.pain.0000000000002510
- Lee, H. J., Choi, E. J., Nahm, F. S., Yoon, I. Y., and Lee, P. B. (2018). Prevalence of unrecognized depression in patients with chronic pain without a history of psychiatric diseases. *Korean J. Pain* 31 (2), 116–124. doi:10.3344/kjp.2018.31.2.116
- Løge-Hagen, J., Sæle, A., Juhl, C., Bech, P., Stenager, E., and Mellentin, A. (2019). Prevalence of depressive disorder among patients with fibromyalgia: Systematic review and meta-analysis. *J. Affect. Disord.* 245, 1098–1105. doi:10.1016/j.jad.2018.12.001
- Markman, J. D., and Dworkin, R. H. (2006). Ion channel targets and treatment efficacy in neuropathic pain. *J. Pain* 7 (1), S38–S47. doi:10.1016/j.jpain.2005.09.008
- Morrison, F. G., Miller, M. W., Wolf, E. J., Logue, M. W., Maniates, H., Kwasnik, D., et al. (2019). Reduced interleukin 1A gene expression in the dorsolateral prefrontal cortex of individuals with PTSD and depression. *Neurosci. Lett.* 692, 204–209. doi:10.1016/j.neulet.2018.10.027
- Nalepa, J., and Kawulok, M. (2019). Selecting training sets for support vector machines: A review. *Artif. Intell. Rev.* 52 (2), 857–900. doi:10.1007/s10462-017-9611-1
- Nilius, B., and Owsianik, G. (2011). The transient receptor potential family of ion channels. *Genome Biol.* 12 (3), 218. doi:10.1186/gb-2011-12-3-218
- Pomares, F. B., Roy, S., Funck, T., Feier, N. A., Thiel, A., Fitzcharles, M.-A., et al. (2020). Upregulation of cortical GABAA receptor concentration in fibromyalgia. *Pain* 161 (1), 74–82. doi:10.1097/j.pain.0000000000001707
- Queiroz, L. P. (2013). Worldwide epidemiology of fibromyalgia. *Curr. pain headache Rep.* 17 (8), 356–6. doi:10.1007/s11916-013-0356-5
- Sacchet, M. D., Prasad, G., Folland-Ross, L. C., Thompson, P. M., and Gotlib, I. H. (2015). Support vector machine classification of major depressive disorder using diffusion-weighted neuroimaging and graph theory. *Front. psychiatry* 6, 21. doi:10.3389/psyt.2015.00021
- Santos, D. d. M., Lage, L. V., Jabur, E. K., Kaziyama, H. H. S., Iosifescu, D. V., Lucia, M. C. S. d., et al. (2011). The association of major depressive episode and personality traits in patients with fibromyalgia. *Clinics* 66 (6), 973–978. doi:10.1590/s1807-59322011000600009
- Sarchielli, P., Mancini, M. L., Floridi, A., Coppola, F., Rossi, C., Nardi, K., et al. (2007). Increased levels of neurotrophins are not specific for chronic migraine: Evidence from primary fibromyalgia syndrome. *J. Pain* 8 (9), 737–745. doi:10.1016/j.jpain.2007.05.002
- Scholkopf, B., Sung, K.-K., Burges, C. J., Girosi, F., Niyogi, P., Poggio, T., et al. (1997). Comparing support vector machines with Gaussian kernels to radial basis function classifiers. *Signal Process. IEEE Trans.* 45 (11), 2758–2765. doi:10.1109/78.650102
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13 (11), 2498–2504. doi:10.1101/gr.1239303
- Siracusa, R., Paola, R. D., Cuzzocrea, S., and Impellizzeri, D. (2021). Fibromyalgia: Pathogenesis, mechanisms, diagnosis and treatment options update. *Int. J. Mol. Sci.* 22 (8), 3891. doi:10.3390/ijms22083891
- Wang, F., Chan, L. W., Cho, W., Tang, P., Yu, J., Shyu, C.-R., et al. (2014). Novel approach for coexpression analysis of E2F1-3 and MYC target genes in chronic myelogenous leukemia. *BioMed Res. Int.* 2014, 439840. doi:10.1155/2014/439840
- Wang, F., Chan, L. W., Tsui, N. B., Wong, S., Siu, P. M., Yip, S. P., et al. (2015). Coexpression pattern analysis of NPM1-associated genes in chronic myelogenous leukemia. *BioMed Res. Int.* 2015, 610595. doi:10.1155/2015/610595
- Wang, F., Meng, F., and Wang, L. (2016). Co-expression pattern analysis of miR-17-92 target genes in chronic myelogenous leukemia. *Front. Genet.* 7, 167. doi:10.3389/fgene.2016.00167
- White, J. P., Cibelli, M., Urban, L., Nilius, B., McGeown, J. G., and Nagy, I. (2016). TRPV4: Molecular conductor of a diverse orchestra. *Physiol. Rev.* 96 (3), 911–973. doi:10.1152/physrev.00016.2015
- Wilke, W. S., Gota, C. E., and Muzina, D. J. (2010). Fibromyalgia and bipolar disorder: A potential problem? *Bipolar Disord.* 12 (5), 514–520. doi:10.1111/j.1399-5618.2010.00848.x
- Wolfe, F., Ross, K., Anderson, J., Russell, I. J., and Hebert, L. (1995). The prevalence and characteristics of fibromyalgia in the general population. *Arthritis & Rheumatism Official J. Am. Coll. Rheumatology* 38 (1), 19–28. doi:10.1002/art.1780380104
- Wong, J. W. (2013). "Signal-to-Noise ratio (SNR)," in *Encyclopedia of radiation oncology*. Editors L. W. Brady and T. E. Yaeger (Berlin, Heidelberg: Springer). doi:10.1007/978-3-540-85516-3\_424
- Wu, C.-T., Dillon, D. G., Hsu, H.-C., Huang, S., Barrick, E., and Liu, Y.-H. (2018). Depression detection using relative EEG power induced by emotionally positive images and a conformal kernel support vector machine. *Appl. Sci.* 8 (8), 1244. doi:10.3390/app8081244
- Yan, W.-J., Ruan, Q.-N., and Jiang, K. (2022). Challenges for artificial intelligence in recognizing mental disorders. *Diagnostics* 13 (1), 2. doi:10.3390/diagnostics13010002
- Zhdanov, A., Atluri, S., Wong, W., Vaghei, Y., Daskalakis, Z. J., Blumberger, D. M., et al. (2020). Use of machine learning for predicting escitalopram treatment outcome from electroencephalography recordings in adult patients with depression. *JAMA Netw. open* 3 (1), e1918377. doi:10.1001/jamanetworkopen.2019.18377
- Zheng, H., Zheng, P., Zhao, L., Jia, J., Tang, S., Xu, P., et al. (2017). Predictive diagnosis of major depression using NMR-based metabolomics and least-squares support vector machine. *Clin. Chim. acta* 464, 223–227. doi:10.1016/j.cca.2016.11.039