EasyChair Preprint

№ 10721

Automatic Detection of Multiple Sclerosis Using Genomic Expression

Abdullah Ahmad, Marwa Hadhuod and Vidan Ghoneim

August 15, 2023

Automatic Detection of Multiple Sclerosis Using Genomic Expression

Abdullah DH.Ahmed[1] [0009-0008-8842-8672],Marwa M.A Hadhoud[2] [0000-0002-2601-3649] ,

Vidan F .Ghoneim[3] [0000-0001-9334-0925]

Biomedical Engineering Department, Faculty of Engineering, Helwan university, Cairo, Egypt

Abstract. Multiple sclerosis (MS) is the prevailing demyelinating disease and the leading cause of neurological disability in the young adult population. Recent microarray gene expression profiling studies have identified a number of genetic variants that contribute to the complex etiology of multiple sclerosis (MS). This study presents a comprehensive analysis of multiple sclerosis (MS) data using microarray technology and machine learning approaches. The goal was to develop a blood biomarker prediction model for the diagnosis of MS. Two experiments were conducted: the first involving Principal Component Analysis (PCA) as a dimension reduction method, and the second utilizing various feature selection techniques.in addition, online STRING database was used for prediction the gene interaction and functional annotation. In the first experiment, PCA was employed with LDA, SVM, and KNN classifiers optimized with different kernel functions. The best accuracy achieved was 95.83% with LDA using 26 components. SVM and KNN classifiers yielded accuracies of 91.67% and 87.5%, respectively. The second experiment focused on feature selection methods (Fisher score, chi-square, relief, and MRMR) combined with LDA, SVM, and KNN classifiers. The best results were obtained with the relief feature selection method, achieving 100% accuracy with KNN using 38 DEG. Fisher score, chi-square and MRMR methods showed higher accuracies of 91.6% ,87.5 and 87.5%, respectively. Functional annotation indicates that these 38 DEG associated with immune and neurological functions. Furthermore, the analysis result suggested that MIF, PTGES3, CYLD and JAK1 may play central roles in gene expression in the pathogenesis of MS.

Keywords: Microarray Technology, Blood Biomarker, Machine Learning, Dimension Reduction.

1 Introduction

Multiple sclerosis (MS) is a chronic neurodegenerative disease primarily affecting young adults, causing plaque formation in the central nervous system (CNS) and demyelination of the myelin sheath surrounding the axon [1]. This disrupts nerve cell function, leading to various neurological issues such as double vision, muscle weakness, motor limitations, and psychological problems [2].

The prevalence of MS varies worldwide, with higher rates in North America and Europe compared to Eastern Asia and sub-Saharan Africa [3]. Globally, around 2.8 mil-

lion people were diagnosed with MS in 2020 [4]. Women are affected about twice as often as men [5-6], and the condition typically worsens between ages 20 and 40 [7].
The exact cause of MS is still largely unknown, but genetic vulnerability and environmental factors, including viral infections and low vitamin D levels due to inadequate sun exposure, may contribute to its development [8, 9].
Diagnosing MS requires a detailed patient history, MRI for white matter lesions, electrophysiological evaluation, and cerebrospinal fluid analysis [10]. Blood transcriptome analysis using DNA microarray technology has been used in research to understand gene expression patterns [11-14].
MS can manifest in four main forms: Primary progressive MS (PPMS), Secondary progressive MS (SPMS), relapsing-remitting MS (RRMS), and clinically isolated syndrome (CIS) [15]. PPMS shows continuous disease progression, while SPMS develops after initial relapsing attacks. RRMS involves clearly defined acute attacks followed by recovery, and CIS is the first episode of neurological symptoms.

2 Theory and Objectives

2.1 Background

The utilization of gene expression analysis has become a valuable method for exploring the underlying molecular mechanisms of many diseases, such as multiple sclerosis (MS). The examination of gene expression profiles enables the detection of genes that are expressed at varying levels, potentially playing a role in the development, progression, and treatment response of diseases. In addition, investigations into gene expression provide an opportunity to identify distinctive patterns that can function as diagnostic biomarkers, hence improving the precision and selectivity of multiple sclerosis (MS) detection.The fundamental principle underlying gene expression analysis is the measurement and quantification of transcript levels for a multitude of genes in a simultaneous manner. The utilization of technology like as microarrays and next-generation sequencing enables the attainment of this objective. These methodologies facilitate the investigation of intricate transcriptional alterations that arise in reaction to illness circumstances, thereby providing insights into plausible molecular pathways and regulatory networks that underlie the development of multiple sclerosis.

Previous MS research has utilized statistical and machine learning techniques to identify disease-related biomarkers.

Studies have employed various biomarkers such as MicroRNA, Gene expression in PBMC ,whole blood microarray analysis, Polymorphisms (SNPs) analysis, and Autoantibodies protein microarrays . For example, researchers have identified microRNAs that are up or downregulated in patients with relapsing-remitting MS (RRMS) compared to healthy controls, which can serve as biomarkers for the disease. Using machine learning techniques, including support vector machines [16] .in addition, Machine learning-based approaches (multilayer perceptron neural network) have been used to predict responsiveness to interferon therapy in MS patients By testing genes in

the interferon signalling pathway for single nucleotide polymorphisms (SNPs). They assessed SNPs' connection with interferon therapy response using automatic relevance determination and backwards elimination [17]. Moreover, peripheral blood mononuclear cells (PBMCs) examine gene expression in MS subtypes Compared to healthy controls, the authors sought gene expression patterns linked with RRMS, SPMS, and PPMS. Affymetrix arrays, ANOVA, PCA, and clustering were used to analyze gene expression data from PBMC subsets like B cells, CD8+ T cells, and monocytes to find potential changes in gene expression. Differences in gene expression, which could be traced to B cells, CD8+ T cells and monocytes, were found between MS patients and HCs, but only minor differences were observed between MS subgroups[18]. Gene expression data are used to create a reliable RRMS diagnostic signature. The researchers used logistic regression with elastic net regression to identify RRMS samples from controls. The study examined peripheral blood mononuclear cell transcriptomes. They tested classifier performance using two feature extraction strategies: one utilizing genes and another using gene pathway data. The two different strategies produced little differences in performance when comparing the 10-fold cross-validation of the training data and prediction on the test data[19]. Gene expression data are used to construct a robust classification model for gene selection. PBMC gene expression profiles from diseased patients and controls are examined. Recursive Feature Elimination (RFE), ROC analysis, and the Boruta algorithm were used to accomplish this. These techniques were used to identify potential genes for the disease of interest; They found an overlapped collection of 8 genes that showed differential expression between MS and control groups[20]. Human protein microarrays were utilized to find MS autoantibody biomarkers in blood samples. Random Forest examined these biomarkers. The study investigated sera from RRMS, SPMS, PD disease, breast cancer, and healthy controls to determine if biomarkers could distinguish MS patients from normal controls and breast cancer patients. Autoantibody biomarkers differentiated MS patients from controls and breast cancer patients, indicating diagnostic potential. But These biomarkers didn't distinguish MS from Parkinson's[21].

2.2 Wider objectives

The aim of this study is to create a statistical model using machine learning techniques for diagnosing MS disease, this objective can be highlighted as create biomarker classification model. Principal Component Analysis (PCA) and Feature selection were applied to identify meaningful patterns in the data. Then, a classifier was developed using supervised learning, using only a subset of the available features.

3 Methodology

The study consisted of two experiments. In the first experiment, Principal Component Analysis (PCA) was employed for dimension reduction[22], and three classifiers (SVM, KNN, and LDA) with different kernel functions were used for microarray data classification. The second experiment focused on identifying

biomarkers for multiple sclerosis disease using feature selection techniques. Four feature selection filter methods (Chi Square[23], MRMR[24], Fisher Score[25], and relief algorithm[26]) were combined with three classifiers (LDA, SVM, and KNN) for this analysis.

3.1 Materials

Using microarray analysis, the expression of many different genes may be analyzed. The proposed analysis is conducted on a dataset of Affymetrix Human Exon 1.0 ST Array presenting N=120 samples (comparing 60 patients with MS and 60 control without MS) and p =18725probesets id(features).

3.2 Experimental Design

In this study The general scheme in the process of classification of microarray data for the detection of proposed MS can be conducted via five stages. Figure 1 shows a flow diagram of these stages.

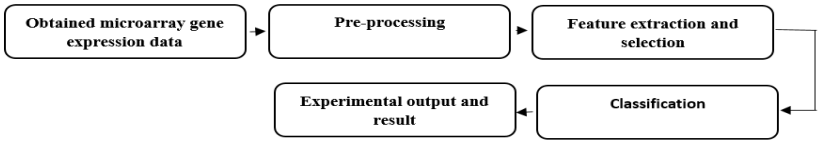

Fig. 1 The proposed cad system is entirely implemented in the mat lab tool except for pre-processing phase, which was implemented using the R tool.

3.3 Data Collection

Gene expression profiles for a total 120 subjects were obtained from the NCBI Database (https ://www.ncbi.nlm.nih.gov/geo/) under accession number of GSE41850 Accordingly, global gene expression in whole blood tissue samples was assessed in 60 multiple sclerosis patients and 60 for control at the time of enrollment (baseline). The specifications of the data can be seen in Table 1.

Table1 . Dataset Specification

Geo dataset	Sample type	Platform	Controls	MS	Visit	Tissue
GSE41850	RNA	GPL16209	60	60	Baseline	Whole blood

3.4 Preprocessing and filtration

- Preprocessing is the procedure used to make the data more usable. We downloaded Series Matrix File, which contains Processed data, and read it.

- Filtration: The filtration procedure is crucial to the analysis of microarray data. Filtering assists in reducing noise, enhancing the quality of the data, and concentrating the research on genes that are more likely to be biologically significant. In the preprocessing of microarray data, the functions genevar, geneentropy, and lowvalfilter in MATLAB play important roles in filtering the data . Therefore we used these three types of filters. Depending on the threshold, each type cleans the data at a certain rate.
- Microarray data were processed using the R package "aroma. Affymetrix". The data were background correct (RMAbackgroundCorrection) and quantum normalized[27]. We performed the summarization step by converting the probe set identifier into a gene official, considering if the probe corresponds to values For multiple sequences, we take the average value and remove probes without gene names if present. Figure 2 summarizes the pre-processing steps.

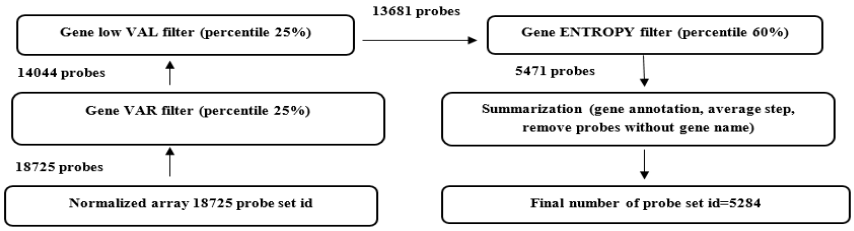

Fig. 2. Pre-processing steps

3.5 Feature extraction (Experiment 1)

The main goal of this stage is to apply dimensionality reduction techniques to explore patterns in the data. From all the possible scenarios, a specific supervised setting is selected. Methods like Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Multidimensional Scaling are utilized to transform the original features into a new feature set, aiming to reveal more meaningful insights. In this study, Principal Component Analysis (PCA) was employed for dimensionality reduction. The general scheme proposed in this experiment involves a process of several stages. Figure 3 shows a flow diagram of these stages.

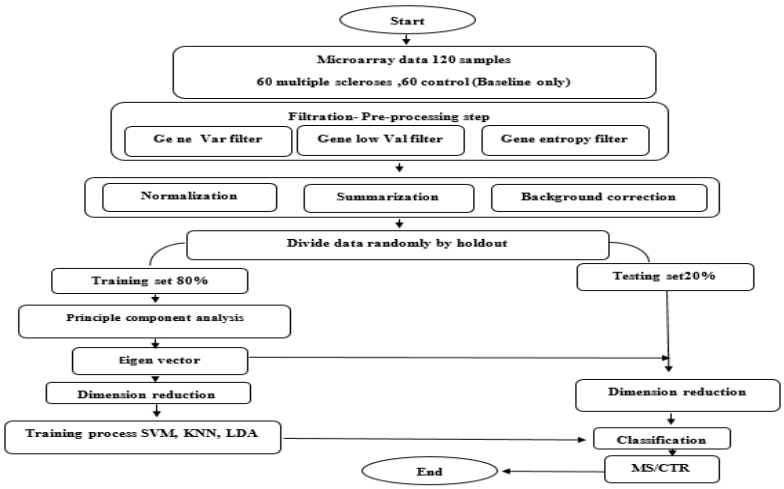

Fig. 3: General scheme of the multiple sclerosis detection process based on principle component analysis PCA

3.6 Feature Selection (Experiment 2)

Feature selection is a crucial process that involves selecting the most relevant and informative features from the initial feature set based on specific criteria. It has gained significant attention in machine learning and pattern recognition as a dimension reduction technique. The advantages of feature selection include gaining a deeper understanding of features, reducing computational requirements, and improving classifier performance. During the selection phase, important features related to the system's target (output) are chosen.

There are various ways to categorize feature selection techniques, with filters, wrappers, embedded, and hybrids being the most common categories. In this experiment, we focus on four feature selection filter methods: Chi Square, Minimum Redundancy, and Maximum Relevance, Fisher score, and Relief Algorithm. The experiment involves multiple stages, and a flow diagram outlining these stages is depicted in Figure 4.

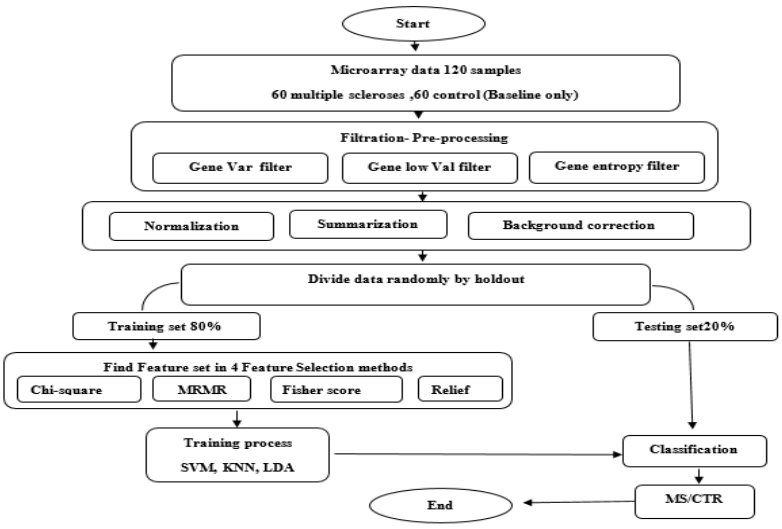

Fig. 4: General scheme of the multiple sclerosis detection process based on feature selection method

4 Results and Discussion

This study conducted two experiments. In the first experiment, microarray data from MS patients and healthy individuals were obtained from the GEO database. (PCA) was used for dimension reduction, and three classifiers (SVM, KNN, and LDA) with different kernel functions were employed for data classification. The best

accuracy was achieved with linear LDA using 26 components, preserving 96.9% variance. Figure 5 presents the outcomes of applying Principal Component Analysis (PCA) as a feature reduction method to the dataset, followed by evaluating the performance of three distinct classifiers: Linear Discriminant Analysis (LDA), Support Vector Machine (SVM) with Radial Basis Function (RBF) kernel, and K-Nearest Neighbors (KNN) using the Euclidian distance metric. The PCA is performed with varying numbers of components, resulting in different levels of variance retained. As the number of components increases, the percentage of variance retained also increases, implying that more relevant information from the original data is preserved in the reduced feature set. The accuracy of the classifiers generally improves with an increasing number of components, suggesting that they benefit from more informative features. However, for some classifiers, accuracy may peak and slightly decline when including too many components, indicating that not all components contribute significant discriminatory information. LDA exhibits relatively higher accuracy compared to SVM and KNN.

The KNN classifier's accuracy fluctuates with varying numbers of components, indicating the need to strike a balance to avoid overfitting and achieve optimal generalization performance. In summary, the results emphasize the importance of selecting an appropriate number of components to achieve efficient dimensionality reduction without sacrificing classification accuracy for different classifiers and datasets. Table 2. Shows the first step results.

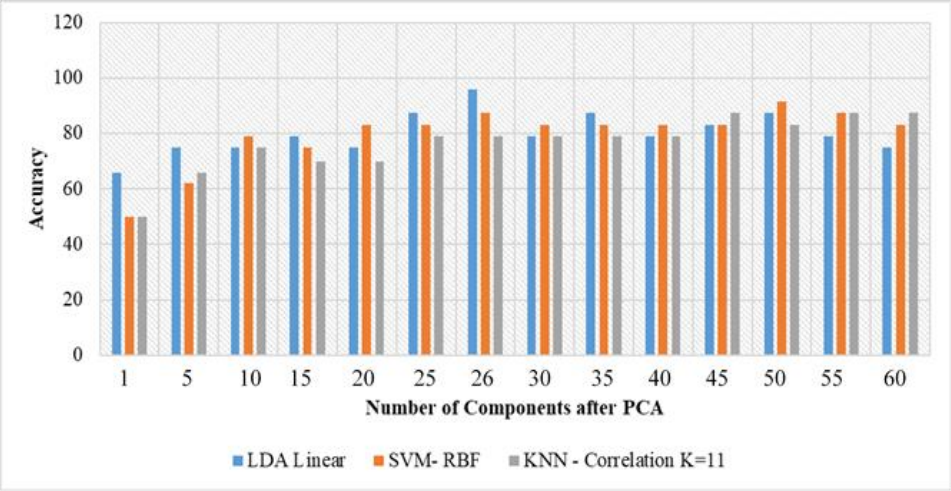

Fig 5. Result of tests for PCA threshold value on system accuracy using the LDA, SVM and KNN classifier

Table 2. Classification results for PCA experiment using multiple sclerosis dataset

Algorithms	PCA no.	Accuracy	Precision	sensitivity	specificity	F-Score
LDA	26	95.83	100	91.67	100	95.65
SVM	50	91.67	91.67	91.67	91.67	91.67
KNN	45	87.5	84.6	91.67	83.3	88

In this step, three classification algorithms, (LDA), (SVM), and (KNN), were evaluated using (PCA) for dimensionality reduction. LDA with liner kernel achieved the highest accuracy (95.83%) and perfect precision (100%), indicating accurate positive predictions. SVM with radial basis function kernel) and KNN with Euclidian distance showed comparable accuracy (91.67% and 87.5% respectively), with SVM slightly better in precision. LDA also demonstrated excellent sensitivity (91.67%) and specificity (100%), while KNN exhibited lower specificity (83.3%). LDA's balanced F-Score (95.65%) highlighted its superior overall performance. The selection of PCA components may have influenced classifier outcomes. Further investigation and optimization of PCA components could enhance the classifiers' effectiveness.

In the second experiment, feature selection techniques were used to identify potential biomarkers for MS disease. Four feature selection algorithms (chi-square, fisher score, MRMR, and relief algorithm) were combined with three classifiers (SVM, KNN, and LDA) for classification. The KNN classifier(correlation distance with k=11) and relief feature selection method k=5) showed the highest predictive accuracy of around 100%, identifying 38 genes associated with MS. According to [28] and [29] the performance of KNN classifier depends significantly on the distance used and k-Nearest Neighbor Parameters and the results showed large gaps between the performances of different distances and k.

Figure 6 presents the outcomes of employing the Relief feature selection method in conjunction with the K-nearest neighbor (KNN) classifier, using different numbers of selected features. The Relief algorithm evaluates the significance of each feature in distinguishing between different classes, while the KNN classifier assigns labels to unseen data points based on their nearest neighbors in the feature space. Figure 6 displays accuracy percentages for various feature subsets, ranging from 1 to 60 features. The results reveal that as the number of selected features increases, the accuracy of the KNN classifier generally improves until reaching a peak at 38 to 44 features, with accuracy values of 100%. after this peak, including more features leads to a slight decline in accuracy, suggesting that some features may not contribute relevant information or could introduce noise. Moreover, the accuracy drops significantly when using 50 or 60 features, indicating potential overfitting. These findings underscore the importance of feature selection in enhancing the performance of machine learning models, highlighting the significance of striking a balance between the number of features selected and the classifier's performance to avoid overfitting and achieve optimal accuracy on unseen data. Table 3 shows a Comparison of classification performances (accuracy) on MS data set.

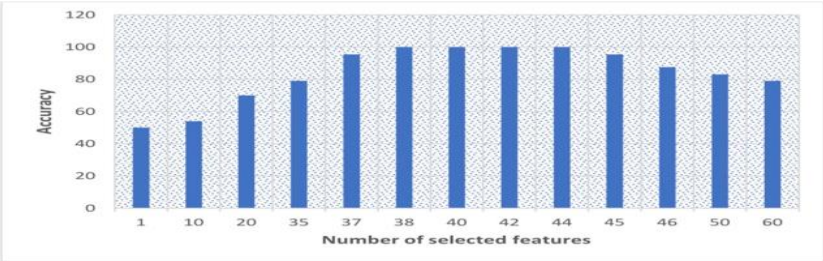

Fig 6. Results for KNN for different Number of Features

Table 3. Comparison of classification performances (accuracy) on MS data set. The number of features used to achieve the maximum is shown inside parenthesis.

Feature Selection Methods	Support Vector Machine	k-Nearest Neighbor	Discriminative
Fisher score	79 (35)	79(20)	91.6(135)
Chi square	87.5(47)	87.5(7)	79(45)
Relief	87.5(30)	100(38)	87.5(17)
MRMR	83(80)	75(27)	87.5(75)

In phase two of the feature selection process, four different feature selection methods were explored: Fisher score, chi-square, relief, and MRMR. The results revealed varying performance for each method in terms of accuracy and the number of selected features for the(SVM) (KNN) and (LDA) classifiers. Fisher score achieved a higher accuracy for LDA but required a larger number of features compared to SVM and KNN. Chi-square demonstrated higher accuracy for KNN and demanded fewer features for its classification. Relief achieved higher accuracy for KNN with a moderate number of selected features. MRMR, on the other hand, resulted in a higher accuracy for LDA while requiring more features. These findings emphasize the importance of selecting an appropriate feature selection method based on the characteristics of the dataset and the specific classifier used, as it significantly impacts the performance of the classification model. Further analysis and comparison of these methods could provide valuable insights for making informed decisions in feature selection for this study.

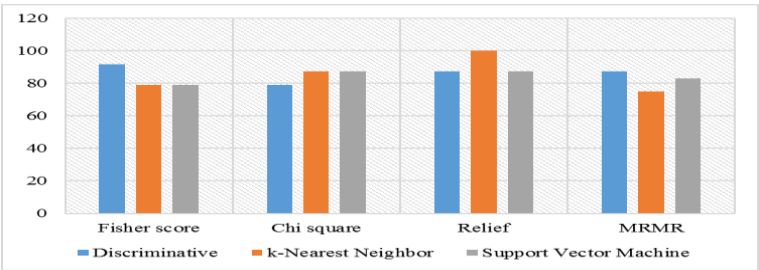

Fig 7: Phase Two Results Comparison

5 Gene Function Analysis

In order to functional annotation and interaction lists of genes in our result, we used the online STRING database (https://string-db.org/) . The retrieved information from the String database was analyzed to annotate the functions of the genes in the list. Protein-protein interactions, co-expression patterns, shared pathways, and predicted functional partners were considered for functional annotation. Gene interactions are shown in Figure 8.

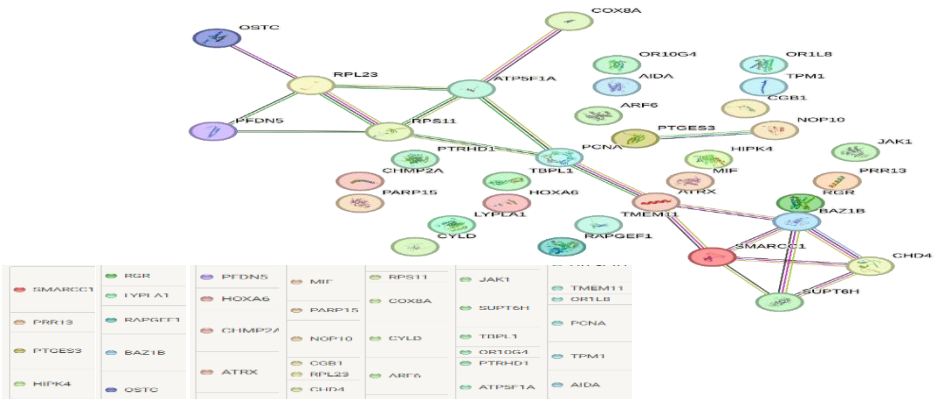

Fig 8: interaction prediction using string database for interaction between list of genes

Table 4 presented provides a complete overview of different genes and their corresponding roles. These genes are involved in various essential biological activities, including but not limited to chromatin remodeling, transcriptional control, immunological modulation, and cell signaling. The table presents succinct information regarding the function of each gene, its participation in distinct pathways, and its significance in diverse physiological settings. This compilation provides a concise yet instructive overview of the genes mentioned in the text, demonstrating their importance in cellular and molecular mechanisms.

Table 4. predicted genes from string database with its functions

Gene name	Function and Involvement
SMARCC1	Subunit of SWI/SNF chromatin remodeling complex
PRR13	Negatively regulates TSP1 expression
HIPK4	Phosphorylates TP53, transcriptional corepressor
RGR	Retinal G protein-coupled receptor, isomerization
LYPLA1	Acyl-protein thioesterase, depalmitoylating activity
RAPGEF1	Guanine nucleotide exchange factor, activates RAS

BAZ1B	Atypical tyrosine-protein kinase, chromatin remodeling
OSTC	Oligosaccharyltransferase complex subunit
PFDN5	Prefoldin subunit, transfers proteins to chaperonin
HOXA6	Homeobox protein, transcriptional regulation
CHMP2A	Component of endosomal sorting complex
ATRX	Regulates chromatin, DNA replication, G4 DNA
PARP15	Negative regulator of transcription, ADP-ribosyltransferase
NOP10	Part of H/ACA ribonucleoprotein complex
CGB1	Chorionic gonadotropin subunit beta 1
RPL23	Ribosomal protein L23
CHD4	Component of histone deacetylase NuRD complex
RPS11	Ribosomal protein, part of uS17 family
COX8A	Cytochrome c oxidase subunit 8A, mitochondrial
ARF6	GTP-binding protein, involved in protein trafficking
SUPT6H	Transcription elongation factor, mRNA processing
TBPL1	TATA box-binding protein-like protein 1
OR10G4	Olfactory receptor, part of G-protein coupled receptor family
PTRHD1	Peptidyl-tRNA hydrolase domain containing 1
ATP5F1A	Mitochondrial ATP synthase alpha subunit
TMEM11	Plays a role in mitochondrial morphogenesis
OR1L8	Olfactory receptor, part of G-protein coupled receptor family
PCNA	Helps DNA polymerase delta replicate efficiently
Protein	Function and Involvement
TPM1	Tropomyosin alpha-1 chain, regulates muscle contraction
AIDA	Impedes JNK activation, embryogenesis pathway
MIF	Regulates immune response, inflammation, CNS function
PTGES3	Synthesizes prostaglandin E2, modulates immune response
CYLD	Regulates cell signaling, inflammation, tumor suppression
JAK1	Part of JAK-STAT pathway, transmits cytokine signals

Also, there was another discovered Genes has more effective on the MS disease, so we will focus on them. These genes are:

- MIF: The cytokine Macrophage Migration Inhibitory Factor (MIF) has been extensively studied for its role in the immune system, central nervous system, and myelin sheath. MIF affects immune response through the immune system. This regulates immunological and inflammatory processes. Macrophage migration inhibitory factor (MIF) can increase inflammatory cytokines, activate immune cells including macrophages and T cells, and activate various immunological processes.

Its role in inflammation and immunological modulation suggests it could affect immune system function. MIF expression in the CNS has been linked to a variety of neurological diseases. This factor can activate microglia and astrocytes. These glial cells maintain CNS homeostasis and respond to damage. MIF regulates CNS inflammation and cellular activation, which may affect neurological processes and the body's response to injury or disease. Myelin protects nerve fibers and improves nerve signal transmission. In conditions like multiple sclerosis (MS), MIF's ability to regulate inflammation and cellular responses may affect myelin's health.

- PTGES3, referred to as prostaglandin E synthase 3, is an enzymatic entity that participates in the process of synthesizing prostaglandin E2 (PGE2), which is classified as a lipid signaling molecule. Prostaglandins, such as PGE2, exhibit a wide range of physiological functions inside the human body, encompassing processes such as inflammation, immune response modulation, and neurological regulation.
- CYLD: The primary function of CYLD is to regulate signaling pathways associated with cell proliferation, inflammation, and tumor suppression. However, it also plays a significant role in the regulation of the immune system. There exists a strong interconnection between the immune system and the central nervous system, whereby abnormalities in immunological signaling can have indirect effects on neurological functions.
- JAK1: a protein in the JAK-STAT signaling system, helps transmit signals from various cytokines and growth factors. The route regulates immune responses, cellular proliferation, and differentiation. Cytokines like interferons and interleukins help immune cells communicate and regulate. The JAK-STAT pathway is involved in immunological responses and cellular development, particularly in the central nervous system and myelin sheath. However, recent research suggest that cytokines and immunological responses may indirectly affect myelin and the central nervous system. Inflammatory cytokines may affect neuronal function, causing neuro inflammation. Multiple sclerosis is affected by this condition.

6 Conclusions

A total of 38 DEG between MS patients and healthy controls were identified. Functional annotation indicated that these DEG were associated with immune and neurological functions these genes may be direct or indirect effect on

immune and nervous system. Furthermore, our analysis result suggested that MIF, PTGES3, CYLD and JAK1 may play central roles in gene expression in the pathogenesis of MS.

References

1. Ropper, A. H., Samuels, M. A. and Klein, J. P. 2014, Multiple sclerosis and allied demyelinative disease, The McGraw Hill Companies, 915.

2. C.-Y. Xia, J.-K. Xu, C.-H. Pan, W.-W. Lian, Y. Yan, B.-Z. Ma, J. He, and W.-K. Zhang, ``Connexins in oligodendrocytes and astrocytes: Pos sible factors for demyelination in multiple sclerosis," Neurochemistry Int., vol. 136, Jun. 2020, Art. no. 104731, doi: 10.1016/j.neuint.2020.104731.
3. Leray, E., Moreau, T., Fromont, A. & Edan, G. Epidemiology of Multiple Sclerosis. Neuroepidemiology 172, 3–13 (2016).
4. Mapping Multiple Sclerosis around the World Key Epidemiology Findings, Atlas of MS, 3rd ed.; The Multiple Sclerosis International Federation (MSIF): London, UK, 2020; Available online: www.atlasofms.org (accessed on 1 December 2020).
5. A. D. Sadovnick and P. A. Baird, "Sex ratio in offspring of patients with multiple sclerosis," The New England Journal of Medicine, vol. 306, no. 18, pp. 1114–1115, 1982.
6. S. V. Ramagopalan, I. M. Yee, D. A. Dyment et al., "Parent of- origin effect in multiple sclerosis: observations from interracial matings," Neurology, vol. 73, no. 8, pp. 602–605, 2009.
7. J. F. Kurtzke, W. F. Page, F.M.Murphy, and J. E.Norman Jr., "Epidemiology of multiple sclerosis in US veterans. 4. Age at onset," Neuroepidemiology, vol. 11, no. 4–6, pp. 226–235, 1992.
8. Smolders J (2011) Vitamin D and multiple sclerosis: correlation, causality, and controversy. Autoimmune Dis 2011:629538. doi: 10.4061/2011/629538
9. van der Mei IA, Ponsonby AL, Dwyer T, Blizzard L, Simmons R, Taylor BV, Butzkueven H, Kilpatrick T (2003) Past exposure to sun, skin phenotype, and risk of multiple sclerosis: case–control study. BMJ 327(7410):316. doi:10.1136/bmj.327.7410.316327/ 7410/316
10. Birnbaum, G., 2006. Making the diagnosis of multiple sclerosis. Adv. Neurol. 98, 111 –124.
11. Nagasaka, Y., Dillner, K., Ebise, H., Teramoto, R., Nakagawa, H., Lilius, L.,Axelman, K., Forsell, C., Ito, A., Winblad, B. et al. (2005) A unique gene expression signature discriminates familial Alzheimer's disease mutation carriers from their wild-type siblings. Proc. Natl Acad. Sci. USA, 102,14854–14859.
12. Tang, Y., Xu, H., Du, X., Lit, L., Walker, W., Lu, A., Ran, R., Gregg, J.P., Reilly, M., Pancioli, A. et al. (2006) Gene expression in blood changes rapidly in neutrophils and monocytes after ischemic stroke in humans: amicroarray study. J. Cereb. Blood Flow Metab., 26, 1089–1102.
13. Pereira, E., Tamia-Ferreira, M.C., Cardoso, R.S., Mello, S.S., Sakamoto-Hojo, E.T., Passos, G.A. and Donadi, E.A. (2004) Immunosuppressive therapy modulates T lymphocyte gene expression in patients with systemic lupus erythematosus. Immunology, 113, 99–105.
14. Schmidt, S., Rainer, J., Riml, S., Ploner, C., Jesacher, S., Achmuller, C.,Presul, E., Skvortsov, S., Crazzolara, R., Fiegl,M. et al. (2006) Identificatio of glucocorticoid-response genes in children with acute lymphoblastic leukemia. Blood, 107, 2061–2069.
15. Lublin, Fred D., and Stephen C. Reingold. "Defining the clinical course of multiple sclerosis: results of an international survey." Neurology 46.4 (1996): 907-911.

16. Keller, A., Leidinger, P., Lange, J., Borries, A., Schroers, H., Scheffler, M., ... & Meese, E. (2009). Multiple sclerosis: microRNA expression profiles accurately differentiate patients with relapsing-remitting disease from healthy controls. *PloS one*, *4*(10), e7440.
17. Calcagno, G., Staiano, A., Fortunato, G., Brescia-Morra, V., Salvatore, E., Liguori, R., ... & Sacchetti, L. (2010). A multilayer perceptron neural network-based approach for the identification of responsiveness to interferon therapy in multiple sclerosis patients. *Information sciences*, *180*(21), 4153-4163.
18. Ratzer, R., Søndergaard, H. B., Christensen, J. R., Börnsen, L., Borup, R., Sørensen, P. S., & Sellebjerg, F. (2013). Gene expression analysis of relapsing–remitting, primary progressive and secondary progressive multiple sclerosis. *Multiple Sclerosis Journal*, *19*(14), 1841-1848.
19. Zhao, C., Deshwar, A. G., & Morris, Q. (2013). Relapsing-remitting multiple sclerosis classification using elastic net logistic regression on gene expression data. *Systems Biomedicine*, *1*(4), 247-253.
20. Guo, P., Zhang, Q., Zhu, Z., Huang, Z., & Li, K. (2014). Mining gene expression data of multiple sclerosis. *PloS one*, *9*(6), e100052.
21. DeMarshall, C., Goldwaser, E. L., Sarkar, A., Godsey, G. A., Acharya, N. K., Thayasivam, U., ... & Nagele, R. G. (2017). Autoantibodies as diagnostic biomarkers for the detection and subtyping of multiple sclerosis. *Journal of neuroimmunology*, *309*, 51-57.
22. Astuti, W., Adiwijaya, 2018. Support vector machineand principal component analysis for microarraydata classification. J. Physics: Conference Series,971: 012003.
23. Ikram, S.T., Cherukuri, A.K. (2016). Intrusion Detection Model Using Fusion of Chi-Square Feature Selection and Multi Class SVM. Journal of King Saudi Saud University –Computer and Information Science. doi: 10.1016/j.jksuci.2015.12.004
24. Rachburee, Nachirat, and Wattana Punlumjeak. "A comparison of feature selection approach between greedy, IG-ratio, Chi-square, and mRMR in educational mining." *2015 7th international conference on information technology and electrical engineering (ICITEE)*. IEEE, 2015.
25. D. Wu, S.Z. Guo, An improved Fisher Score feature selection method and its application, Chinese Journal of Liaoning Technical University 38 (5) (2019) 472–479.
26. K. Kira and L. Rendell, "The feature selection problem: traditional methods and a new algorithm," in Proceedings of AAAI-92, 1992.
27. Nickles, Dorothee, et al. "Blood RNA profiling in a large cohort of multiple sclerosis patients and healthy controls." Human molecular genetics 22.20 (2013): 4194-4205.
28. Abu Alfeilat, Haneen Arafat, et al. "Effects of distance measure choice on k-nearest neighbor classifier performance: a review." *Big data* 7.4 (2019): 221-248.
29. Astuti, Widi, and Adiwijaya. "Support vector machine and principal component analysis for microarray data classification." Journal of Physics: Conference Series. Vol. 971. IOP Publishing, 2018.