Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 1051388, 10 pages
https://doi.org/10.1155/2022/1051388

Research Article

Supervised Machine Learning Empowered Multifactorial Genetic Inheritance Disorder Prediction

Taher M. Ghazal,[1,2] Hussam Al Hamadi,[3] Muhammad Umar Nasir,[4] Atta-ur-Rahman,[5] Mohammed Gollapalli,[6] Muhammad Zubair,[7] Muhammad Adnan Khan,[8] and Chan Yeob Yeun[3]

[1]*School of Information Technology, Skyline University College, Sharjah 1797, UAE*
[2]*Network and Communication Technology Lab, Center for Cyber Security, Faculty of Information Science and Technology, Universiti Kebangsaan, Malaysia 43600, Malaysia*
[3]*Center for Cyber Physical Systems, Khalifa University, Abu Dhabi, 127788, UAE*
[4]*Riphah School of Computing & Innovation, Faculty of Computing, Riphah International University Lahore Campus, Lahore 54000, Pakistan*
[5]*Department of Computer Science, College of Computer Science and Information Technology, Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam 31441, Saudi Arabia*
[6]*Department of Computer Information Systems, College of Computer Science and Information Technology, Imam Abdulrahman Bin Faisal University, P.O. Box 1982, Dammam 31441, Saudi Arabia*
[7]*Faculty of Computing, Riphah International University, Islamabad 45000, Pakistan*
[8]*Pattern Recognition and Machine Learning Lab, Department of Software, Gachon University, Seongnam, Gyeonggido 13120, Republic of Korea*

Correspondence should be addressed to Hussam Al Hamadi; hussam.alhamadi@ku.ac.ae and Muhammad Adnan Khan; adnan@gachon.ac.kr

Received 10 March 2022; Revised 15 April 2022; Accepted 3 May 2022; Published 31 May 2022

Academic Editor: Paolo Gastaldo

Fatal diseases like cancer, dementia, and diabetes are very dangerous. This leads to fear of death if these are not diagnosed at early stages. Computer science uses biomedical studies to diagnose cancer, dementia, and diabetes. With the advancement of machine learning, there are various techniques which are accessible to predict and prognosis these diseases based on different datasets. These datasets varied (image datasets and CSV datasets) around the world. So, there is a need for some machine learning classifiers to predict cancer, dementia, and diabetes in a human. In this paper, we used a multifactorial genetic inheritance disorder dataset to predict cancer, dementia, and diabetes. Several studies used different machine learning classifiers to predict cancer, dementia, and diabetes separately with the help of different types of datasets. So, in this paper, multiclass classification proposed methodology used support vector machine (SVM) and K-nearest neighbor (KNN) machine learning techniques to predict three diseases and compared these techniques based on accuracy. Simulation results have shown that the proposed model of SVM and KNN for prediction of dementia, cancer, and diabetes from multifactorial genetic inheritance disorder achieved 92.8% and 92.5%, 92.8% and 91.2% accuracy during training and testing, respectively. So, it is observed that proposed SVM-based dementia, cancer, and diabetes from multifactorial genetic inheritance disorder prediction (MGIDP) give attractive results as compared with the proposed model of KNN. The application of the proposed model helps to prognosis and prediction of cancer, dementia, and diabetes before time and plays a vital role to minimize the death ratio around the world.

1. Introduction

Dementia, a degenerative brain illness, is a significant health issue in terms of global health, public health, and population health [1]. Understanding the development of dementia illness and aiding early identification of dementia have been recent focal points of study in the disciplines of neuroimaging and genetics [2]. Numerous genome-wide association studies have also been undertaken since 2007 to discover genetic variations such as single nucleotide polymorphisms that are linked to dementia [3]. Researchers have continued to make significant discoveries in the interdisciplinary disciplines of machine learning, neuroimaging, genomics, and dementia diagnosis and prediction as artificial intelligence tools have improved [4, 5]. Recent advancements in artificial intelligence (AI) technology, notably machine learning approaches, have demonstrated their usefulness in health-related and genetic medicine applications [6, 7]. Under specific environmental settings, biological characteristics are the consequence of interactions between gene sequences and gene interactions.

The machine learning model is appropriate for investigating the link between these variables and phenotype. Through machine learning of genomes, genome machine learning (GML) investigates the connection among genetic variants and characteristics. Although genome-wide association study (GWAS) is utilized to detect correlations between single nucleotide variants and cancer, it depends on linkage analysis to find sick genes and necessitates more intimate segregated locations [8, 9]. Diabetes mellitus is a chronic illness characterized by persistent hyperglycemia induced by a variety of factors. The primary cause is a deficiency in insulin secretion. Typical symptoms include polyuria, polydipsia, polyphagia, and weight loss, which may be accompanied by skin itching. Long-term carbohydrate, lipid, and protein metabolism problems can also result in several chronic consequences, including chronic progressive illness, hypofunction, and tissue failure, as well as organs such as the eyes, kidneys, nerves, heart, and blood vessels.

Large quantities of data hide important information and insights in the age of big data. A significant quantity of data filtered by relevant data sources is merged into a data set for data mining to predict dementia, cancer, and diabetes. Following that, users may use machine learning algorithms to classify and analyze this dataset. This not only allows patients to avoid and cure dementia, cancer, and diabetes at an early stage through prediction but it also saves a significant amount of time and money. This paper employs a variety of algorithms to train an integrated data set before proposing an algorithm that may utilize the medical history of an early genetic problem to predict dementia, cancer, and diabetes. Major purpose of this study is to get efficient prediction results for cancer, dementia, and diabetes using different machine learning algorithms and test machine learning model performance using numerous statistical parameters. With the help of improvised results, medical field will get major benefits from this study and play their pivot role in serving for people.

2. Related Work

In recent years, two large multicenter studies have been conducted to discover biomarkers for early diagnosis of dementia and the development of Medical Council of India (MCI) to dementia: the auxiliary nursing midwifery, which is located in Europe, and the dementia disease neuroimaging initiative, which is based in the United States. Furthermore, at national center for biotechnology information gene expression omnibus, a substantial quantity of publicly available gene expression data on dementia has been given [10]. As a result, numerous research studies, particularly gene expression-based investigations, have been published to identify the informative genes linked to dementia. The brain net study examined 113 well-characterized postmortem brain tissue samples, resulting in the identification of 21 dysregulated genes in dementia [11]. A study of 87 brain tissue samples by Liang et al. found that the genes encoding the subunits of mitochondrial components were substantially lower expressed in the brain tissues of dementia patients. Xu et al. [12] discovered that an early change in protein1 might. Cause dementia by examining the ribonucleic acid expression of brain tissues from dementia patients. Although numerous studies utilizing gene expression data have found significant patterns, the majority of the gene expression data was collected from biopsies or autopsy-based and patients data samples, which makes extrapolation to clinical settings challenging. Only a few research [13] utilized blood expression data to identify important genes associated with dementia or to predict early dementia. Cooper et al. [13] presented research that included 186 dementia patients and 204 controls from three different data sets, indicating that progranulin expression levels in the blood are higher in dementia. Tae-WooKim et al. [14] created an operational lung cancer decision tree. The occupational safety and health researchers institute recorded 153 instances of lung cancer between 1992 and 2007. The goal was to evaluate if the condition was acknowledged as lung cancer connected to age, gender, years of smoking, histology, industry size, delay, work hours, and exposure of independent factors. The characterization and relapse test concept are utilized along the route of word-related cell degradation markers in the lungs. The greatest signal of the lungs cancer detection model was its introduction to well-known lung disease experts. In 2014, Maciej Ziba et al. [15] presented powered SVM, which is devoted to addressing imbalanced outcomes. For unequal data, the suggested approach coupled the benefits of employing set classifiers with cost-sensitive support vectors. A technique for extracting choices from the improved SVM was also provided. The effectiveness of the suggested method was then assessed by comparing the performance of the imbalanced data with that of other algorithms. Finally, in lung cancer patients, enhanced SVM was utilized to estimate life expectancy following surgery. Numerous techniques, including classic machine learning methods [16], such as support vector machine, decision tree (DT), logistic regression, and others, have recently been applied to predict diabetes. The authors in [17] proposed the linear discriminant analysis

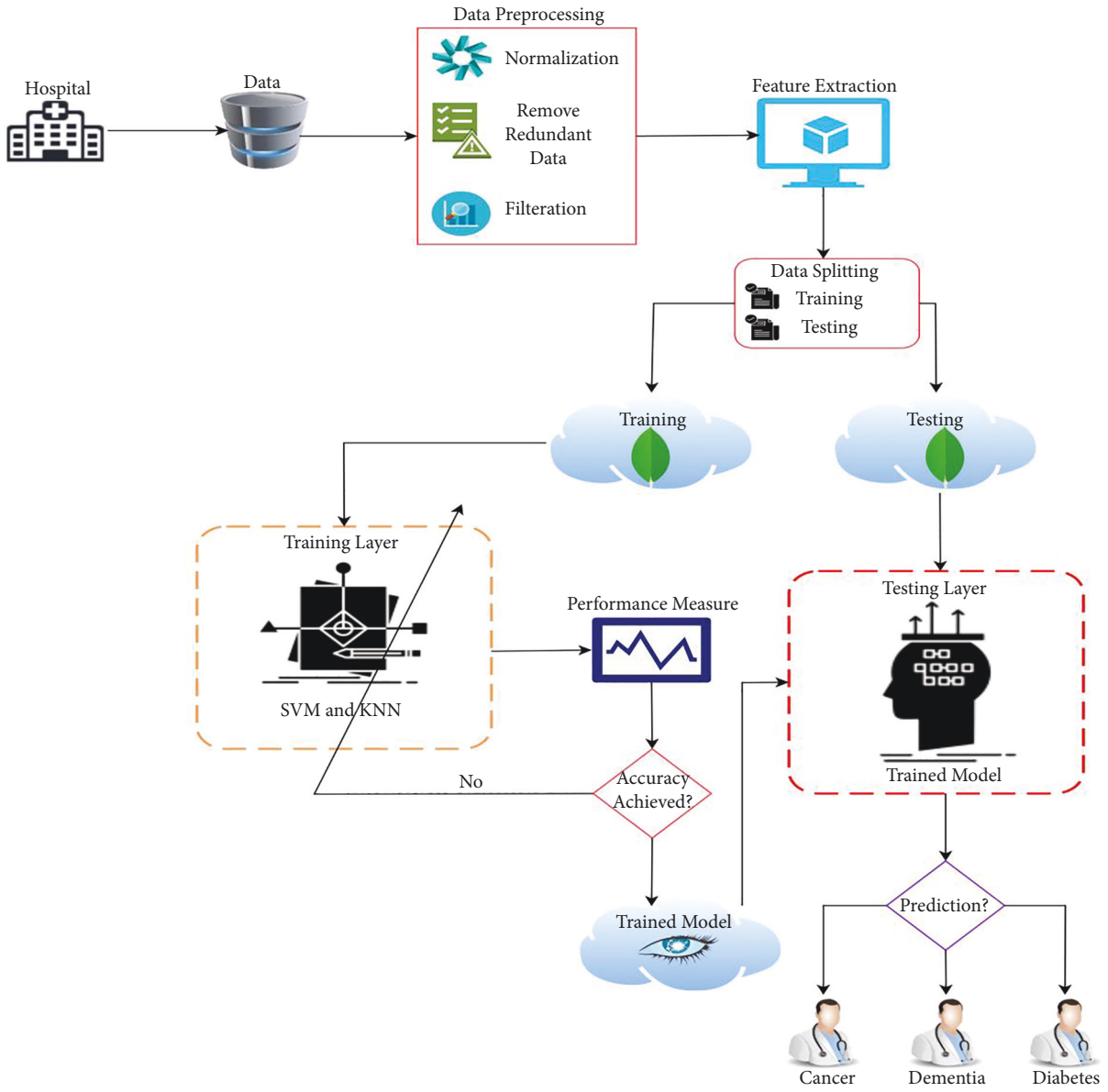

Figure 1: Proposed machine learning-based cancer, dementia, and diabetes prediction model from multifactorial genetic disorder.

diabetes prediction method. The authors utilized linear discriminant analysis in this system to decrease dimensionality and extract features. To deal with the high-dimensional data sets, the authors in [18] built logistic regression-based prediction models for several type 2 diabetes prediction beginnings. The authors in [19] focused on hyperglycemia and employed support vector regression (SVR), a regression analysis issue, to predict diabetes. Furthermore, joint techniques are being used in an increasing number of research to increase accuracy [16]. The authors in [20] presented rotation forest, a novel ensemble technique that incorporates 30 machine learning algorithms. The authors in [21] suggested a machine learning technique that altered the SVM prediction criteria. In 2017, Kee Pang Soh et al. [22] created a cancer prediction model using SVM and applied their model on sequenced tumored DNA and achieved 77.7% prediction accuracy. In 2019, Javier De Velasco Oriol et al. [23] suggested a machine learning technique for dementia prediction using neuroimaging and their machine learning model achieved 75% classification accuracy. In 2013, BAssam Farran et al. [24] used four machine learning techniques such as linear regression, SVM, KNN, and multifactorial dimension reduction to predict diabetes, and their model achieved 81.3% highest prediction accuracy from SVM. In [25], researcher proposed feature extraction techniques with machine learning algorithms for the prediction of dementia. They measured the performance of their model with the help of different statistical parameters. In this study [25], researchers classify the dementia using different machine learning algorithms and achieved 83% testing accuracy. All previous researches used different limited data for single class classification with limited approaches of machine learning algorithms and feature extraction techniques and achieved lowest prediction accuracy.

Machine learning algorithms are frequently utilized to forecast diabetes, dementia, and cancer and achieve better outcomes. SVM and KNN are two prominent machine learning algorithms in the medical area, and they have high sorting power. SVM and KNN are two newly prominent machine learning methods that outperform others in many ways. In this paper, to predict diabetes, dementia, and cancer, the proposed model utilized a support vector machine and K-nearest neighbors using multifactorial genetic inheritance disorder data of sick patients and predicted multiclass diseases efficiently.

3. Research Methodology

The proposed machine learning-based cancer, dementia, and diabetes prediction model from multifactorial genetic disorder is shown in Figure 1. In the first phase, data are collected from hospital and stored in database; right after this step, the proposed model performed two steps; first is data preprocessing to normalize data using different normalization techniques and removing the duplicate data using different queries and in second step the proposed model performed correlation technique to extract high performed features for further training and testing step. So, the proposed model divided data into testing and training and stored in separate data clouds. In the second phase, the proposed model used machine learning algorithms to train data and check the model performance if the performance of trained achieve the benchmark than the trained model stored in cloud database otherwise retrain the models. In the final and third phase, the proposed model imported testing data from test data cloud and trained model from train cloud and performed testing queries to predict the cancer, dementia, and diabetes.

4. Dataset

The dataset in this research paper is obtained from open-source Kaggle [26]. This dataset consists of 2067 medical patients' history which is diagnosed with cancer, dementia, and diabetes. The total dataset consists of 32 independent attributes including, patient age, genes in mothers' side, inherited from father, maternal gene, no. of previous abortions, and so on and one dependent attribute.

Table 1 shows some of the dataset's attributes descriptions, 1 indicates "yes" and 2 indicates "no" and in gender attribute 3 indicates "ambiguous" gender.

Figure 2 shows the no. of patients from the targeted class and targeted class patients 1822, 152, and 93 from diabetes, dementia, and cancer, respectively.

5. Support Vector Machine

Support vector machine is a general linear predictor that uses supervised learning to perform three-class data classifications. Its decision limit for handling learning patterns is the maximum margin hyperplane [23]. Support vector machine calculates empirical risk using the pivot loss function and optimizes structural risk by including a regularization component in the solution system. It is a scarcity and robustness predictor [24].

The kernel technique, which is one of the most prevalent kernel learning methods, allows the support vector machine to execute a nonlinear sort.

SVM hypothesis is as follows:

$$F_\theta(q) = \begin{cases} 1 & \text{if } \theta^v f >= 0 \\ 0 & \text{otherwise} \end{cases}, \theta^v f = \theta_0 f_0 + \theta_1 f_1 + \cdots + \theta_z f_z. \tag{1}$$

SVM loss function is as follows:

$$D(\theta) = k\left[\sum_{g=1}^{z} y^{(i)} \text{Cost}_1\left(\theta^v\left(f^{(g)}\right) + \left(1 - y^{(g)}\right)\text{Cost}_0\left(\theta^v\left(f^{(g)}\right)\right)\right]. \tag{2}$$

SVM regularized loss function is as follows:

$$D(\theta) = C\left[\sum_{g=1}^{z}\left[y^{(g)}\text{Cost}_1\left(\theta^r\left(f^{(g)}\right) + \left(1 - y^{(g)}\right)\text{Cost}_0\left(\theta^T\left(f^{(g)}\right)\right)\right] + \frac{1}{2}\sum_{h=1}^{z}\theta_h^2\right. \tag{3}$$

6. Simulation Results

In this study, MATLAB R2021 is used for simulation and predictions. The proposed model is used to train and test patients' data using machine learning techniques like SVM and KNN. In starting of the simulation proposed model, split the dataset (2067 instances) into 70% (1447 instances) for training and 30% (620 instances) for testing. After applying the proposed research model on the training dataset using both SVM and KNN machine learning algorithms, we get the SVM trained model and KNN trained model for prediction purposes. After the availability of the trained model, these trained models were used to predict cancer, dementia, and diabetes using the testing dataset. After that, we select the best prediction model by applying different statistical performance parameters like classification accuracy (CA), missclassification rate (MCR), sensitivity, specificity, F1-score, positive predicted value (PPV), negative predicted value (NPV), false positive ratio (FPR), false negative ratio (FNR), likelihood positive ratio (LPR), and likelihood negative ratio (LNR) on simulation results. Simulation results of the prediction proposed model are elaborated below with respect to the confusion matrix and different performance statistical parameters. In confusion

matrix, ∂ represents true positive results, μ represents true negative results, $\emptyset$ represents false positive results, and Ω represents false negative results.

$$\partial_i = \frac{\varphi_i}{\varrho_i}, \tag{4}$$

where φ is for predicted class and ϱ for true class:

$$\begin{aligned}
\mu_i &= \sum_{j=1}^{3}\left(\frac{\varphi_i}{\varrho_{j\neq i}}\right),\\
\emptyset_i &= \sum_{j=1}^{3}\left(\frac{\varphi_{j\neq i}}{\varrho_i}\right),\\
\Omega_i &= \sum_{j=1}^{3}\left(\frac{\varphi_{j\neq i}}{\varrho_{j\neq i}}\right),\\
\mathrm{CA} &= \frac{\varphi_i/\varrho_i + \sum_{j=1}^{3}\left(\varphi_i/\varrho_{j\neq i}\right)}{\varphi_i/\varrho_i + \sum_{j=1}^{3}\left(\varphi_i/\varrho_{j\neq i}\right) + \sum_{j=1}^{3}\left(\varphi_{j\neq i}/\varrho_i\right) + \sum_{j=1}^{3}\left(\varphi_{j\neq i}/\nu_{j\neq i}\right)}, * 100,\\
\mathrm{CMR} &= 100 - \left(\frac{\varphi_i/\varrho_i + \sum_{j=1}^{3}\left(\varphi_i/\varrho_{j\neq i}\right)}{\varphi_i/\varrho_i + \sum_{j=1}^{3}\left(\varphi_i/\varrho_{j\neq i}\right) + \sum_{j=1}^{3}\left(\varphi_{j\neq i}/\varrho_i\right) + \sum_{j=1}^{3}\left(\varphi_{j\neq i}/\varrho_{j\neq i}\right)} * 100\right),\\
\mathrm{Sensitivity} &= \frac{\varphi_i/\varrho_i}{\varphi_i/\varrho_i + \sum_{j=1}^{3}\left(\varphi_{j\neq i}/\varrho_{j\neq i}\right)} * 100,\\
\mathrm{Specifity} &= \frac{\sum_{j=1}^{3}\left(\varphi_i/\varrho_{j\neq i}\right)}{\sum_{j=1}^{3}\left(\varphi_i/\varrho_{j\neq i}\right) + \sum_{j=1}^{3}\left(\varphi_{j\neq i}/\varrho_i\right)} * 100,\\
F1-\mathrm{Score} &= \frac{2\,\varphi_i/\varrho_i}{2\varphi_i/\varrho_i + \sum_{j=1}^{3}\left(\varphi_{j\neq i}/\varrho_i\right) + \sum_{j=1}^{3}\left(\varphi_{j\neq i}/\varrho_{j\neq i}\right)} * 100,\\
\mathrm{PPV} &= \frac{\varphi_i/\varrho_i}{\varphi_i/\varrho_i + \sum_{j=1}^{3}\left(\varphi_{j\neq i}/\varrho_i\right)} * 100,\\
\mathrm{NPV} &= \frac{\sum_{j=1}^{3}\left(\varphi_i/\varrho_{j\neq i}\right)}{\sum_{j=1}^{3}\left(\varphi_i/\varrho_{j\neq i}\right) + \sum_{j=1}^{3}\left(\varphi_{j\neq i}/\varrho_{j\neq i}\right)} * 100,\\
\mathrm{FPR} &= 100 - \left(\frac{\sum_{j=1}^{3}\left(\varphi_i/\varrho_{j\neq i}\right)}{\sum_{j=1}^{3}\left(\varphi_i/\varrho_{j\neq i}\right) + \sum_{j=1}^{3}\left(\varphi_{j\neq i}/\varrho_i\right)} * 100\right),\\
\mathrm{FPR} &= 100 - \left(\frac{\varphi_i/\varrho_i}{\varphi_i/\varrho_i + \sum_{j=1}^{3}\left(\varphi_{j\neq i}/\varrho_{j\neq i}\right)} * 100\right),\\
\mathrm{LPR} &= \frac{\varphi_i/\varrho_i/\varphi_i/\varrho_i + \sum_{j=1}^{3}\left(\varphi_{j\neq i}/\varrho_{j\neq i}\right) * 100}{100 - \left(\sum_{j=1}^{3}\left(\varphi_i/\varrho_{j\neq i}\right)/\sum_{j=1}^{3}\left(\varphi_i/\varrho_{j\neq i}\right) + \sum_{j=1}^{3}\left(\varphi_{j\neq i}/\nu_i\right) * 100\right)},\\
\mathrm{LNR} &= \frac{100 - \left(\varphi_i/\varrho_i/\varphi_i/\varrho_i + \sum_{j=1}^{3}\left(\varphi_{j\neq i}/\varrho_{j\neq i}\right) * 100\right)}{\sum_{j=1}^{3}\left(\varphi_i/\varrho_{j\neq i}\right)/\sum_{j=1}^{3}\left(\varphi_i/\varrho_{j\neq i}\right) + \sum_{j=1}^{3}\left(\varphi_{j\neq i}/\varrho_i\right) * 100}.
\end{aligned} \tag{5}$$

Table 1: Description of dataset attributes.

No.	Attributes	Values
1	Patient age	0–16
2	Genes in mothers' side	1: yes; 2: no
3	Inherited from father	1: yes; 2: no
4	Maternal gene	1: yes; 2: no
5	Paternal gene	1: yes; 2: no
6	Gender	1: male; 2: female; 3: ambiguous
7	Birth asphyxia	1: yes; 2: no

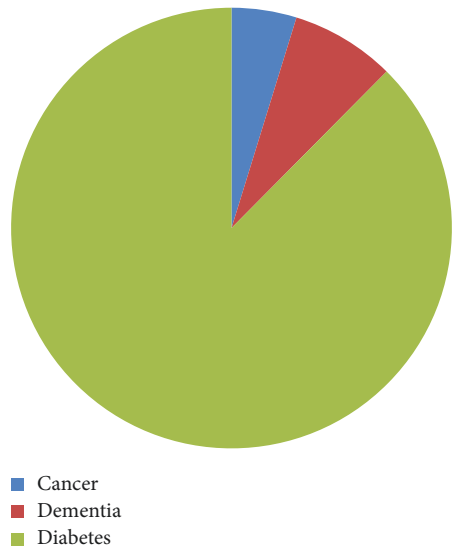

Figure 2: No. of patients from targeted classes.

Table 2: Training performance of the proposed three-class SVM and KNN-based model.

Instances (1447)	SVM			KNN		
	Dementia	Cancer	Diabetes	Dementia	Cancer	Diabetes
Dementia	5	0	98	2	0	101
Cancer	0	57	10	0	56	11
Diabetes	6	2	1268	2	1	1273

Table 3: Testing performance of the proposed three-class SVM and KNN-based model.

Instances (1447)	SVM			KNN		
	Dementia	Cancer	Diabetes	Dementia	Cancer	Diabetes
Dementia	3	0	46	6	0	43
Cancer	0	27	3	0	20	10
Diabetes	0	0	541	11	0	530

Table 2 shows the proposed KNN and support vector machine model-based cancer, dementia, and diabetes prediction from multifactorial genetic inheritance disorder during the training phase. Total 1447 instances were used during training simulation; furthermore, these instances were divided into 103, 67, and 1276 sections of dementia, cancer, and diabetes, respectively. During the training session of the proposed SVM model, predicts 5, 57, 1268, 2, and 10 are correctly classified and 98 and 6 are wrongly classified. Similarly, the proposed KNN-based model training session predicts 2, 56, 1273, 1, and 11 are correctly classified and 101 and 2 are wrongly classified.

Table 3 shows the proposed support vector machine and KNN model-based cancer, dementia, and diabetes prediction from multifactorial genetic inheritance disorder during the testing phase. Total 620 instances were used during the testing phase; furthermore, these instances were divided into 49, 30, and 541 sections of dementia, cancer, and diabetes, respectively. During the testing phase of the proposed SVM model, predicts 3, 27, 541, and 3 are correctly classified and 46 is

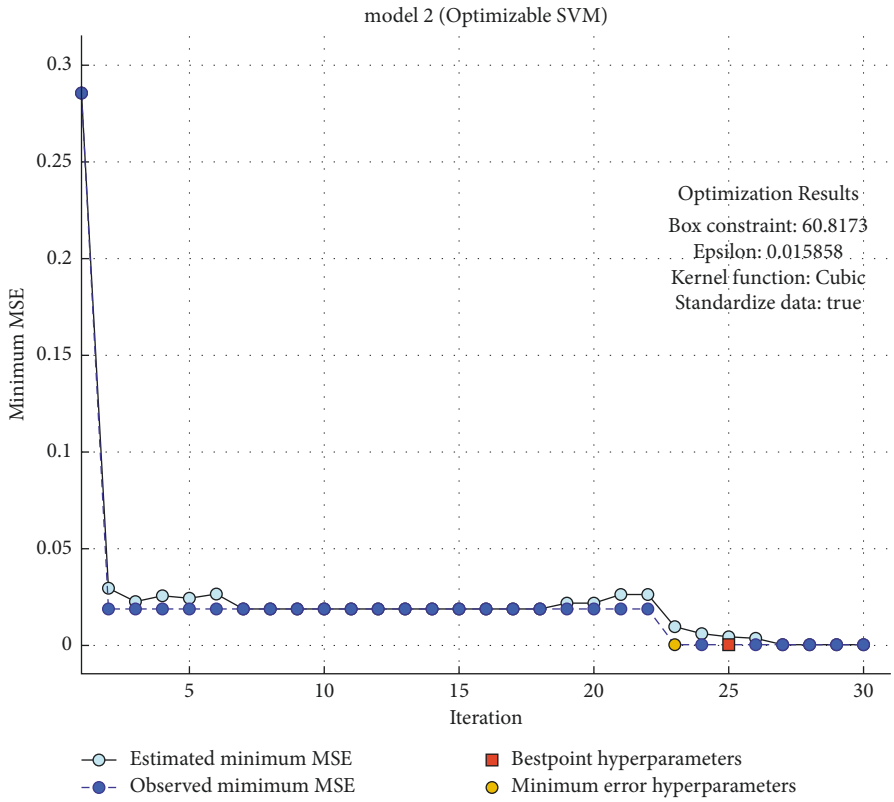

Figure 3: Performance of the proposed SVM-based model w.r.t MSSE vs. iterations.

Table 4: Training simulation results of dementia class by the proposed model.

Instances (1446)	CA (%)	CMR (%)	Sensitivity (%)	Specificity (%)	F1-score (%)	NPV (%)	FPR (%)	FNR (%)	LPR (%)	LNR (%)	PPV (%)
SVM	94.77	5.23	45.45	95.16	11.9	99.5	4.84	54.55	9.39	0.57	6.84
KNN	94.84	5.16	50	92.99	3.73	92.99	7.01	50	7.13	0.53	1.94

Table 5: Training simulation results of cancer class by the proposed model.

Instances (1446)	CA (%)	CMR (%)	Sensitivity (%)	Specificity (%)	F1-score (%)	NPV (%)	FPR (%)	FNR (%)	LPR (%)	LNR (%)	PPV (%)
SVM	99.17	0.83	96.61	99.27	90.47	99.27	0.73	3.39	132.34	0.034	85.07
KNN	99.17	0.83	98.24	99.2	90.32	99.20	0.8	1.76	122.8	0.017	83.58

wrongly classified. Similarly, the proposed KNN-based model testing session predicts 6, 20, 530, and 10 are correctly classified and 43 and 11 are wrongly classified. It is observed that the proposed model of SVM has the highest correctly classified instances as compared with the proposed model of KNN.

Figure 3 shows the performance of the proposed SVM-based model with respect to minimum mean square error (MMSE) vs. iterations. It clearly observed that the proposed support vector machine-based model congregated at the 5th iteration with 0.00482 MMSE.

Table 4 shows the training simulation results of dementia class using different statistical parameters by the proposed model of SVM and KNN. Table 5 shows the training simulation results of cancer class using different statistical parameters by the proposed model of SVM and KNN. Table 6 shows the training simulation results of diabetes class using different statistical parameters by the proposed model of SVM and KNN. Table 7 shows the testing simulation results of dementia class using different statistical parameters by the proposed model of SVM and KNN. Table 8 shows the testing simulation results of cancer class using different statistical parameters by the proposed model of SVM and KNN. Table 9 shows the testing simulation results of diabetes class using different statistical parameters by the proposed model of SVM and KNN.

Table 10 shows the number of performance statistical parameters used to calculate the performance of the proposed SVM and KNN prediction model to predict dementia, cancer, and diabetes from multifactorial genetic inheritance disorder. Table 8 shows the statistical parameter like

Table 6: Training simulation results of diabetes class by the proposed model.

Instances (1446)	CA (%)	CMR (%)	Sensitivity (%)	Specificity (%)	F1-score (%)	NPV (%)	FPR (%)	FNR (%)	LPR (%)	LNR (%)	PPV (%)
SVM	91.97	8.03	92.15	88.57	95.62	36.47	11.43	7.85	8.06	0.088	99.37
KNN	92.04	7.96	91.9	95.08	95.67	34.11	4.92	8.1	18.67	0.088	99.76

Table 7: Testing simulation results of dementia class by the proposed model.

Instances (620)	CA (%)	CMR (%)	Sensitivity (%)	Specificity (%)	F1-score (%)	NPV (%)	FPR (%)	FNR (%)	LPR (%)	LNR (%)	PPV (%)
SVM	92.58	7.42	100	92.54	12.24	100	7.46	0	13.4	0	6.12
KNN	91.29	8.71	35.29	92.86	18.18	98.07	7.14	64.71	4.94	0.69	12.24

Table 8: Testing simulation results of cancer class by the proposed model.

Instances (620)	CA (%)	CMR (%)	Sensitivity (%)	Specificity (%)	F1-score (%)	NPV (%)	FPR (%)	FNR (%)	LPR (%)	LNR (%)	PPV (%)
SVM	99.51	0.49	100	99.4	94.7	100	0.6	0	166.6	0	90
KNN	98.38	1.62	100	98.3	80	100	1.7	0	58.82	0	66.6

Table 9: Testing simulation results of diabetes class by the proposed model.

Instances (1446)	CA (%)	CMR (%)	Sensitivity (%)	Specificity (%)	F1-score (%)	NPV (%)	FPR (%)	FNR (%)	LPR (%)	LNR (%)	PPV (%)
SVM	92.09	7.91	91.69	100	95.66	37.97	0	8.31	0	0.083	100
KNN	89.67	10.33	90.90	70.27	94.30	32.91	29.73	9.1	3.05	0.129	97.96

Table 10: Proposed model parameter results.

Instances (2067)	SVM		KNN	
	Training (%) (1446 instances)	Testing (%) (620 instances)	Training (%) (1446 instances)	Testing (%) (620 instances)
Accuracy	92.8	92.5	92.8	91.2
Miss-rate	7.2	7.5	7.2	8.8

Table 11: Comparative analysis with previous work.

Work	Model	Dataset	Classification accuracy (%)
Kee Pang Soh et al. [20]	Logistic regression, random forest	Cancer mutate data	77.7
Javier De Velasco Oriol et al. [21]	Linear ML	SNP dataset	75
Bassam Farran et al. [22]	Logistic regression, KNN	Kuwait health network data	81.3
Proposed model for prediction of dementia, cancer, and diabetes	Machine learning (SVM and KNN)	Genome disorder data	92.5

accuracy and miss-rate results of proposed SVM and KNN models, so the proposed SVM model achieves 92.8% and 7.2% of training accuracy and miss classification rate, respectively. Similarly, the proposed model of KNN achieves 92.8% and 7.2% of training accuracy and miss classification rate, respectively. In the prediction phase, the proposed SVM model achieves 92.5% and 7.5% of testing accuracy and miss classification rate, respectively, and the proposed KNN model achieves 91.2% and 8.8% of testing accuracy and miss classification rate, respectively.

Table 11 shows the comparative analysis of the proposed model with previous studies. As in Table 11, the proposed model outclassed all mentioned previous studies and achieved highest classification accuracy in all three diseases cancer, dementia, and diabetes as well as the proposed model of SVM for the prediction of dementia, cancer, and diabetes from multifactorial genetic inheritance disorder achieves the highest test classification accuracy as compared with the proposed model of KNN. The proposed model achieved highest prediction accuracy because it used different data preprocessing techniques to

clean up the data and correlation techniques for extraction of highly relatable features, and the proposed model used all these features to predict the cancer, dementia, and diabetes.

7. Conclusion and Future Work

Machine learning plays a bigger role in the classification of different diseases in medical and biomedical fields. In this study, the proposed model used two machine learning techniques SVM and KNN to predict dementia, cancer, and diabetes from multifactorial genetic inheritance disorders. The proposed model analyzed the prediction results with respect to different statistical performance parameters. In this study, the proposed model used patients' multifactorial genetic inheritance disorder history to predict dementia, cancer, and diabetes because patient medical history puts a major impact on prediction results. The proposed model SVM achieves the highest testing prediction classification accuracy of 92.5% as compared with the proposed model of KNN. This study will play a major part in the medical field to early predict these dangerous diseases in the early stages of life with the help of the patient's genetic history and procure these diseases in early stages. Major advantage of this study is to predict the multiclass prediction of genome disorders, i.e., cancer, dementia, and diabetes with the help of major machine learning algorithms and correlation techniques. On the other hand, there will be more improvements in the light of data and other machine learning and transfer learning techniques. Furthermore, in the future, this study will expand to predict all these three diseases with the help of genetic sequence data and also, with the help of mitochondrial genetic inheritance disorder prediction of leigh syndrome and mitochondrial myopathy using machine learning techniques, federated machine learning and transfer learning will play a major role in the genetic field.

Data Availability

The data used in this paper can be obtained from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this work.

References

[1] A. Kumar, A. Singh, and Ekavali, "A review on alzheimer's disease pathophysiology and its management: an update," *Pharmacological Reports*, vol. 67, no. 2, pp. 195–203, 2015.

[2] D. Veitch, M. Weiner, P. Aisen et al., "Understanding disease progression and improving alzheimer's disease clinical trials: recent highlights from the alzheimer's disease neuroimaging initiative," *Alzheimer's and Dementia*, vol. 15, no. 1, pp. 106–152, 2019.

[3] S. Andrews, B. Fulton, and A. Goate, "Interpretation of risk loci from genome-wide association studies of alzheimer's disease," *The Lancet Neurology*, vol. 19, no. 4, pp. 326–335, 2020.

[4] M. Tanveer, B. Richhariya, R. Khan et al., "Machine learning techniques for the diagnosis of alzheimer's disease: a review," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 16, no. 1, pp. 1–35, 2020.

[5] P. Khan, M. Kader, R. Islam et al., "Machine learning and deep learning approaches for brain disease diagnosis: principles and recent advances," *IEEE Access*, vol. 9, Article ID 37622, 2021.

[6] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[7] P. Srinivasu, J. Sivasai, M. Ijaz, A. Bhoi, W. Kim, and J. Kang, "Classification of skin disease using deep learning neural networks with mobilenet v2 and lstm," *Sensors*, vol. 21, no. 8, pp. 2852–2885, 2021.

[8] A. Sud, B. Kinnersley, and R. Houlston, "Genome-wide association studies of cancer: current insights and future perspectives," *Nature Reviews Cancer*, vol. 17, no. 11, pp. 692–704, 2017.

[9] G. Battineni, N. Chintalapudi, and F. Amenta, "Performance analysis of different machine learning algorithms in breast cancer predictions," *EAI Endorsed Transactions on Pervasive Health and Technology*, vol. 6, no. 23, Article ID 166010, 2020.

[10] T. Barrett, D. Troup, S. Wilhite et al., "Ncbi geo: archive for functional genomics data sets—10 years on," *Nucleic Acids Research*, vol. 39, pp. 1005–1010, 2010.

[11] F. Durrenberger, F. Fernando, S. Kashefi et al., "Common mechanisms in neurodegeneration and neuroinflammation: a brainnet europe gene expression microarray study," *Journal of Neural Transmission*, vol. 122, no. 7, pp. 1055–1068, 2015.

[12] M. Xu, D. Zhang, R. Luo et al., "A systematic integrated analysis of brain expression profiles reveals yap1 and other prioritized hub genes as important upstream regulators in alzheimer's disease," *Alzheimer's and Dementia*, vol. 14, no. 2, pp. 215–229, 2018.

[13] A. Cooper, D. Nachun, D. Dokuru et al., "Progranulin levels in blood in alzheimer's disease and mild cognitive impairment," *Annals of clinical and translational neurology*, vol. 5, no. 5, pp. 616–629, 2018.

[14] W. Kim and H. Koh, "Decision tree of occupational lung cancer using classification and regression analysis," *Safety and Health at Work*, vol. 1, no. 2, pp. 140–148, 2010.

[15] M. Zięba, J. Tomczak, M. Lubicz, and J. Świątek, "Boosted svm for extracting rules from imbalanced data in application to prediction of the post-operative life expectancy in the lung cancer patients," *Applied Soft Computing*, vol. 14, no. 3, pp. 99–108, 2014.

[16] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, "Machine learning and data mining methods in diabetes research," *Computational and Structural Biotechnology Journal*, vol. 15, no. 2, pp. 104–116, 2017.

[17] C. Duygu and D. Esin, "An automatic diabetes diagnosis system based on an lda-wavelet support vector machine classifier," *Expert Systems with Applications*, vol. 38, no. 7, pp. 8311–8315, 2011.

[18] N. Razavian, S. Blecker, A. Schmidt, A. Smith-McLallen, S. Nigam, and D. Sontag, "Population- level prediction of type 2 diabetes from claims data and analysis of risk factors," *Big Data*, vol. 3, no. 4, pp. 277–287, 2015.

[19] E. Georga, V. Protopappas, D. Ardigo et al., "Multivariate prediction of subcutaneous glucose concentration in type 1 diabetes patients based on support vector regression," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 1, pp. 71–81, 2013.

[20] A. Ozcift and A. Gulten, "Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms," *Computer Methods and Programs in Biomedicine*, vol. 104, no. 3, pp. 443–451, 2011.

[21] L. Han, S. Luo, J. Yu, L. Pan, and S. Chen, "Rule extraction from support vector machines using ensemble learning approach: an application for diagnosis of diabetes," *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 2, pp. 728–734, 2015.

[22] P. Soh, E. Szczurek, T. Sakoparnig, and N. Beerenwinkel, "Predicting cancer type from tumour dna signatures," *Genome Medicine*, vol. 9, no. 1, pp. 104–123, 2017.

[23] J. Oriol, E. Vallejo, K. Estrada, and J. Tamez, "Benchmarking machine learning models for late-onset alzheimer's disease prediction from genomic data," *BMC Bioinformatics*, vol. 20, pp. 202–218, 2019.

[24] B. Farran, A. Channanath, K. Behbehani, and T. Thanaraj, "Predictive models to assess risk of type 2 diabetes, hypertension and comorbidity: machine-learning algorithms and validation using national health data from Kuwait—a cohort study," *BMJ Open*, vol. 3, no. 5, Article ID e002457, 2013.

[25] C. Kavitha, V. Mani, S. Srividhya, O. Khalaf, and A. Tavera, "Early-stage alzheimer's disease prediction using machine learning models," *Frontiers in Public Health*, vol. 10, Article ID 853294, 2022.

[26] R. Arya, *Of Genome and Genetics*, Kaggle, 2021, https://www.kaggle.com/datasets/aryarishabh/of-genomes-and-genetics-hackerearth-ml-challenge/code.