EXTREME GRADIENT BOOSTING MODEL FOR PREDICTION OF GENOME DISORDER

D.N.V.Pavan Manikanta[1], Dr Sheik Meerasharif[2],

[1 2 3] Bonam Venkata Chalamayya Engineering College (Autonomous), Odalarevu
[1]pavandesineedi46@gmail.com, [2]sharief.mtech@gmail.com

Abstract. Genetic illness prediction is an important and timely issue in the realm of biomedical science. Mutations in the genome are the root cause of many diseases with significant global mortality rates, including Alzheimer's, cancer, diabetes, cystic fibrosis, leigh syndrome, and others. Theoretical and explanatory approaches to predicting genetic abnormalities have been developed through prior research. Genetic data has expanded to practically include the entire genome and protein, and methods based on deep learning and machine learning have been created to forecast genomic abnormalities in response. Concurrently with the introduction of machine learning techniques, deep learning methods also emerged. Studies on the forecasting of genetic anomalies have previously employed a variety of learning strategies, including supervised, unsupervised, and semi-supervised approaches. Most of these studies used genetic sequence data to make predictions about binary dilemmas. These methods produced dubious results since they were less accurate and relied on binary class prediction algorithms, which ignore the pasts of individuals with genetic anomalies. The majority of the approaches relied on RNA gene sequences, which led to frequent issues when dealing with auction data. Here, we use the XGBoost Algorithm to foretell genome multiclass disease from a huge dataset utilising an advanced genome disorder prediction model (AGDPM). AGDPM outperformed the trained XGBoost Algorithm in every category, with an average accuracy of 92.65% in both the training and testing phases of the study. Therefore, the state-of-the-art genome disorder prediction model can reliably predict genome disorder and analyse a large quantity of patient genome disorder data thanks to the incorporation of a multi-class prediction technique. Multiple statistical performance metrics demonstrate that AGDPM may accurately predict diseases caused by a single gene, mitochondrial genes, and multiple genes. As a result, AGDPM will help biomedical researchers manage mortality rates and anticipate genetic disorders.

Keywords: AGDPM, XGBoost, RNA, Deep Learning, DNN, SGID, MGID and CNN

1 Introduction

It is estimated that almost 2,000 different human diseases can be traced back to a single faulty gene, making them monogenic syndromes. The underlying genes for each illness present themselves in somewhat different ways, resulting in a wide range of phenotypic manifestations. Therefore, establishing phenotype-gene correlations is a crucial biological activity that aids researchers and medical professionals in

understanding the fundamental genetic pathways behind disorders. The identification of disease-causing genes aids in patient diagnosis and sheds light on the complex network of genetic interactions. In other words, a possible genetic disease can be detected by studying the causative mutant genotypes during the sickness gene identification procedure. The same way Single nucleotide changes, single nucleotide additions or deletions, complete gene loss, and other genetic anomalies can all have an impact on disease-causing genes. Positional cloning, linkage analysis, and mutation analysis are all examples of time-honored approaches to identifying pathogen genes. First, using linkage analysis on human pedigrees, the susceptible chromosomal interval is discovered, which is roughly where the disease-associated candidate genes are located. The use of positional cloning to sequence a set of putative genes in the region is the second topic covered. This approach incorporates both spatial and transcriptional mapping. A human genetic disorder is an inherited condition manifested from conception due to a genetic or chromosomal abnormality. There are two primary categories of genetic illnesses: single-gene diseases and complicated disorders. One gene aberration caused by a single mutation in the structure of deoxyribonucleic acid is a severe shortcoming. These problems are easily passed down from one generation to the next. Mandolin diseases is a term used to describe this group of illnesses. Complex diseases are the pathological outcome of a confluence of environmental, behavioural, and lifestyle factors, and genetic defects account for only a small fraction of the phenotypes associated with these diseases. A mutation in a single gene is the sole cause of a single gene disorder. The wide variety of single-gene illnesses is due to the fact that they might originate in any gene.All single gene disorders share the same core genetic and psychosocial care needs despite their wide variation in presentation. the ability to make educated decisions about risk management strategies and provide emotional and practical assistance to those who are ill, whether they are young or old. It's associated with alterations in mitochondrial DNA that doesn't come from the nucleus. There are as many as ten circular strands of deoxyribonucleic acid in the mitochondrial genome. After becoming fertilised eggs, they keep their organelles in tact. Therefore, mothers always end up becoming the source of their children's illnesses. The symptoms of mitochondrial disease are lactic acidosis, stroke-like episodes, eye abnormalities, and encephalopathy.Inherited disorders have various root causes. Several diseases have several causes, including gene alterations that work in tandem with dietary and environmental variables. Polygenic illness can also be referred to as complex illness [2]. One complex genetic disorder underlies diabetes, Alzheimer's, and cancer.

2 Objective

The AGDPM used numerous statistical performance parameters to predict the results of the multifactorial gene inheritance disease simulation. Furthermore, genetic illnesses might be multifactorial, which means that genetic factors contribute to the development of only a subset of the phenotypes associated with the disorder. Diseases with multiple causal factors, or risk factors, include those caused by both genetic predisposition and environmental influences. A mutation in a single gene is the sole cause of a single gene disorder. The wide variety of single-gene illnesses is due to the fact that they might

originate in any gene. Despite their clinical distinctions, all single-gene illnesses are inherited, share a common biological basis, and require the same fundamental genetic and counselling services. the ability to make educated decisions about risk management strategies and provide emotional and practical assistance to those who are ill, whether they are young or old. It's associated with alterations in mitochondrial DNA that doesn't come from the nucleus. There are between five and ten circular strands of deoxyribonucleic acid that make up each mitochondrial genome. After becoming fertilised eggs, they keep their organelles in tact. Therefore, mothers always end up becoming the source of their children's illnesses. The symptoms of mitochondrial disease are lactic acidosis, stroke-like episodes, eye abnormalities, and encephalopathy. These diseases, which are frequently the result of interplays between environmental and nutritional factors, may involve many mutations. It's sometimes called complicated illness or polygenic disease. One complex genetic disorder underlies diabetes, Alzheimer's, and cancer. An alternative to conventional methods of genetic prediction is machine learning. Due to advancements in the area, as well as growing data sets and computing power, deep learning has become increasingly popular in recent years. These methods are useful in statistical genetics because they enable the identification of interactions between several loci without the need to assume additivity and because of the high dimensionality with which they operate, making it difficult to predict the relative importance of various factors.

2.1 Problem Statement

In the realm of genetics and medical research, it is essential to forecast genome disorders. Although Deep Neural Networks (DNNs) have shown a lot of promise in tackling this issue, their generalization and performance are constrained by overfitting. Convolutional Neural Networks (CNNs) are limited by the increasing spatial correlation of zeroed-out values in output feature maps. In order to avoid overfitting, dropout is frequently employed.

The current setup recommends using Checkerboard Dropout, a structured dropout method, to improve performance and generality while also tackling the spatial correlation problem. Despite its advantages, the Checkerboard Dropout may still have problems that need fixing.

2.2 Existing System

Dropout is a method employed by contemporary Deep Neural Networks (DNNs) to combat overfitting. During a dropout, features from feature maps are removed at random. However, the dropout's applicability to CNNs is constrained by an increase in the spatial correlation of the zeroed-out values in the output feature maps, which in turn hinders the network's performance and generalisation.
Drop Block, which is an organised dropout used to drop a continuous zone and reduce the unpredictability of the standard dropout, has recently been used efficiently to alleviate the spatial correlation problem.

Disadvantage of Existing System

- The recommends using Checkerboard Dropout as a fix for the overfitting problem.
- An efficient structured dropout technique for mitigating randomness and spatial correlation problems while augmenting model generalisation is the Checkerboard Dropout

2.3 Proposed System

Complex Multiple gene abnormalities can cause a wide variety of symptoms. These include multifactorial genome disorder, mitochondrial gene inheritance disorder, and single gene inheritance disorder.

New advances in genomic technology have made it possible to acquire genetic data with greater precision.Hundreds of people with abnormalities have been found in many large-scale genetic studies, including those for MGD and SGID. Despite the mountain of data our study has produced, pinpointing the specific disease-causing genes has proven challenging. After becoming fertilised eggs, they keep their organelles in tact. Therefore, mothers always end up becoming the source of their children's illnesses.

Advantages of Proposed System

- A gradient descent method is used to minimise the loss when adding new models.
- Its independence in doing feature engineering.
- The suggested model, XGBoost Algorithm, obtained 92.65% prediction accuracy using patients' clinical feature base data.
- The suggested model, which also had ideal space and computational complexity, employed the perfect XGBoost Algorithm to forecast this illness.
- It improved dramatically in terms of result prediction.

3 Related Works

New advances in genomic technology have made it possible to acquire genetic data with greater precision. Hundreds of people with abnormalities have been found in numerous large-scale genetic studies, including those for MGD and SGID [4, 5]. Finding the specific disease-causing genes has been challenging despite the abundance of data from this investigation [6]. The fact that different disturbances within a single disorder module often produce similar phenotypes, as well as the close relationships between proteins and phenomena networks (where genes are appended endpoints if they indicate associated phenotypic states), suggest that genetic information is particularly useful [7]. connection between transcription factor networks and the genome [8]. Furthermore, anomalies observed in interactome distant neighbours create unique phenotypes [6]. There are methods out there for predicting disease based on genes that take into account all of these factors. In this investigation, a binary support vector machine was used to aggregate data from several sources. Binary learning algorithms, both adaptive and maladaptive [10, 11], have been proposed as a means of

sifting through the residual collection in the hopes of discovering previously undiscovered genes or diseases. Recent years have witnessed the successful implementation of deep learning and machine learning in many biological applications. Despite being able to handle massive data sets with substantial noise, complexity, and/or error levels, deep learning and machine learning algorithms only produce a small number of trustworthy estimates of probability distributions and data production processes.

4 Methodology

In order to effectively treat genetic illnesses, early diagnosis is crucial for both clinicians and the biomedical industry. In this investigation, we suggest AGDPM for the early diagnosis of multi-class genetic anomalies. The training model of the XGBoost algorithm and the AGDPM are used to illustrate the investigation's flow. This method will use a streamlit framework to promote user involvement since it anticipates output that includes mitochondrial gene inheritance illnesses, single-gene inheritance disorders, and multifactorial gene inheritance disorders without the need for a physician.

Modules Name:

- Data gathering,
- dataset creation,
- data preparation, model selection,
- analysis, and prediction,
- accuracy on the test set
- saving the trained model

Module Description:

1) Data Collection:

This is the first step in the real process of collecting data and creating a machine learning model. This is an important stage since the amount and quality of data we are able to gather will determine how effectively the model works.

Manual interventions, online scraping, and other techniques are used in data collection.

2) Dataset:

The collection contains 22084 unique bits of data. The 45 columns in the dataset are described in detail below.

1. Patient Id: Patient Id with "Genetic Disorder" written on it.
2. Patient Age: The age of the patient or the user
3. Mother's side genes - maternal genes, whether or not they are present.
4. Inherited from father to father: Parents use DNA to pass on characteristics or traits to their children, such as blood type and eye colour.
5. Maternal gene: Genes that produce or deposit RNA or protein byproducts in the oocyte, or are found in the fertilised egg or embryo prior to the onset of zygotic gene expression, are known as maternal genes.

6. Paternal gene: Paternal inheritance is the term used to describe any characteristic that a father passes on to his offspring.
The measurement of the amount of red, white, and platelet-rich blood in the body is called the blood cell count (mcL).
8. Patient First Name, which is the patient's surname
9. Father's name and family name
10. Name of mother and father
11. Age of mother - age of mother 12. Age of father - age of father
13. Institution Name: The hospital's or institution's name
14. Institute's Location: Hospital or Institutional
15. Status: Is the person or patient still living or has passed away?
16. Respiratory Rate (breaths/min): The brain's respiratory centre controls and determines how quickly people breathe.
17. Heart Rate (rates/min): The frequency of the heartbeat, also called the beats per minute, or bpm, is established by counting the number of heartbeats (also called pulse rate, or heart rate) that transpire each minute.
18. Test 1: Is it finished?
19. Test 2: Is it finished?
20. Test 3: Is it finished?
21. Test 4: Is it finished?
22. Test 5: Is it finished?
23. Parental consent - Also known as parental involvement laws, parental consent laws require one or more parents to provide their assent or notify their child before the child is allowed to legally participate in a particular activity.
24. Check if fellow-up is at a high or low level.
25. Gender: Male, Female, or Indeterminate
26. Birth asphyxia - Asphyxia, also called asphyxiation, is a condition in which breathing irregularities allow the body to get insufficient oxygen. asphyxia during childbirth
27. Autopsy reveals birth defect (if any) - An autopsy, also called an obduction, an autopsiacadaverum, a post-mortem examination, or a necropsy, is a surgical procedure that involves a thorough examination of a corpse through dissection to determine the manner, mode, and cause of death as well as to evaluate any disease or injury that may be present for instructional or research purposes.
28. Place of Birth: The birthplace
29. Information about folic acid (peri-conceptional): Folic acid is a form of vitamin B. It aids the body in producing new, healthy cells.
30. H/O serious maternal disease - Indicates an unanticipated result of labour and delivery that had a major impact on the patient's mother in the short or long term 31. H/O radiation exposure (x-ray) - Indicates whether the patient has ever been exposed to radiation
32. H/O substance abuse - Indicates if a parent has previously struggled with drug addiction.
33. Assisted Conception: IVF/ART - Indicates the kind of infertility therapy
34. Previous pregnancy abnormalities - any history of unexplained things in prior pregnancies Certainly or no
35. Number of prior abortions – total amount of prior abortions

36. Birth defects – Indicates if a patient is afflicted with birth defects.
37. White blood cell count (number of White Blood cells) expressed in thousands per microliter
38. Blood test result: Normal, slightly abnormal, unclear, and abnormal values
39. Symptom 1: Does Symptom 1 exist? 40. Symptom 2: Yes or no 41 for Symptom 2. Symptom 3: Yes or no, symptom 42. Symptom 4: Yes or no, symptom 43. Symptom 5: Yes or no 44 for Symptom 5. Genetic Disorder - Professional doctor detection of genetic disorders
45. Type of Disorder – Subclass

3) Data Preparation:
Compile the data and prepare it for training. Eliminate duplicates, correct errors, deal with missing numbers, normalise, convert data types, and other potential clutter.
By randomising the data, the effects of the particular order in which we collected and/or otherwise processed our data are erased.
Conduct additional exploratory analysis, such as visualising data to find significant class imbalances or relationships between variables (beware of bias!).
separated into sets for assessments and training.

4) Model Selection:
After utilising the XGBoost and Support Vector Machine methods, which produced accuracy of 98% and 80% on the train set, respectively, we developed this method.

5) Analyze and Prediction:
Out of the entire dataset, we only chose two attributes:
1 A description of the health values is given.
2 Outcome: indicates the type of genetic condition that the patient or individual has.

6) Accuracy on test set:
We obtained accuracy of 92.65% & 41.40% on the test set.

7) Saving the Trained Model:
You are ready to put your training to work when you: The first step in deploying your trained and validated model in a production setting is to save it as a.pkl file using a library like Pickle.
Verify that Pickle has been set up properly in your environment.
At this stage, the module will import the model and generate an.pkl file for export.

5 Algorithm Used in Project

XGBOOST Algorithm:
The regularised (L1 and L2) objective function in XgBoost consists of a convex loss function (based on the difference between the predicted and target outputs) and a penalty term for model complexity (i.e., the regression tree functions), both of which must be minimised for the method to be effective. New trees are added to the training

process to predict the errors or residuals from earlier trees, and these trees are blended with the original trees to get the final prediction.

Because of its high performance, scalability, and accuracy, XGBoost is widely utilised in image classification, text mining, and recommender systems applications. Input features used by AGDPM include data on genetic diseases.

6 Data Flow Diagram

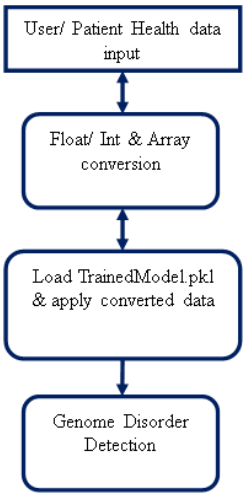


Fig. 1. Data Flow Diagram

7 System Architecture

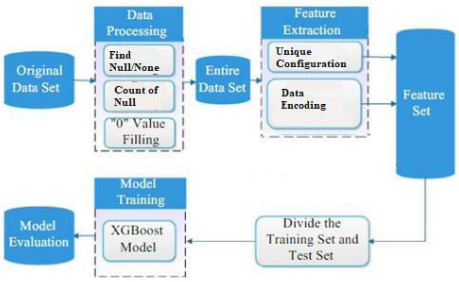
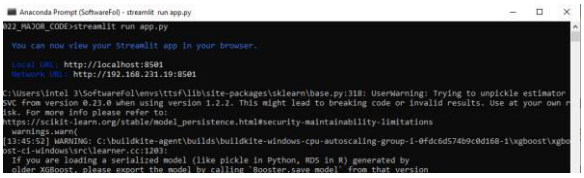
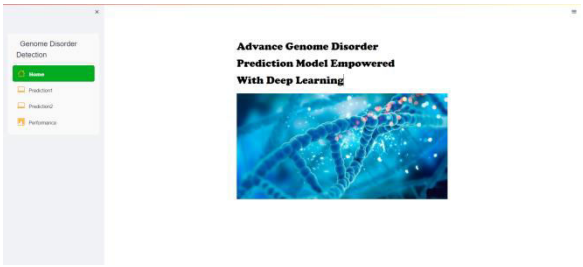
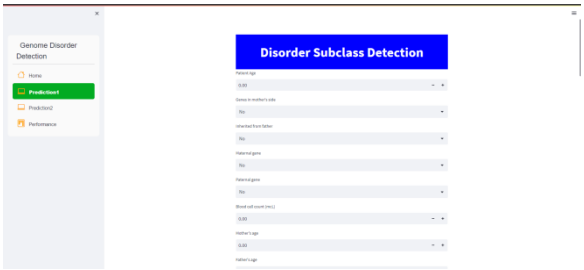


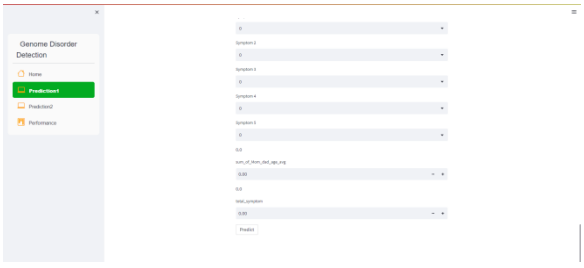
Fig. 2. System Architecture Of Project

8 System Architecture

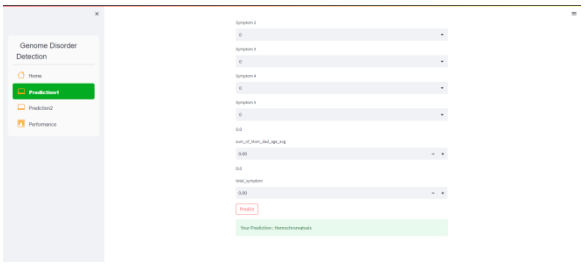


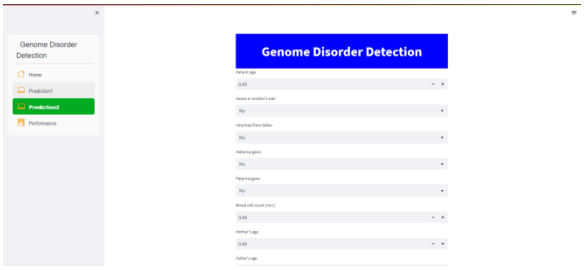


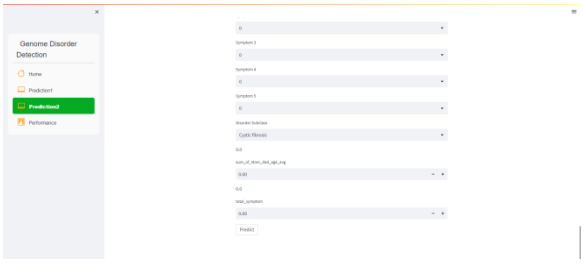


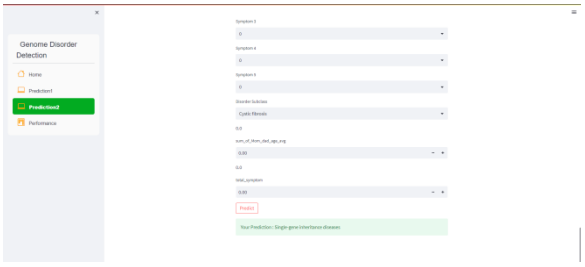


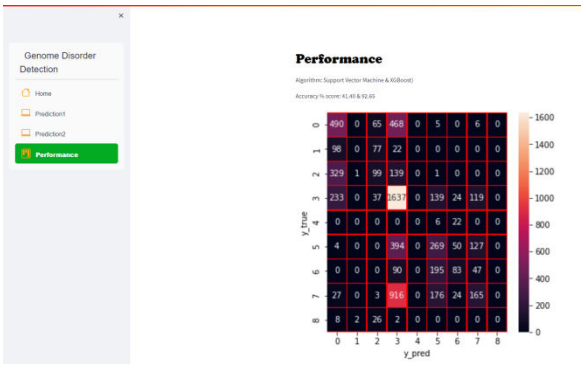
10

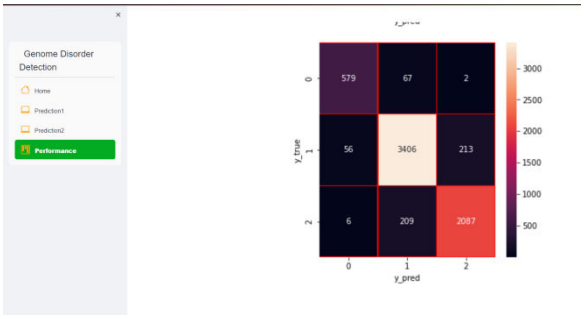

Genome Disorder Detection
Home
Prediction1
Prediction2
Performance


Genome Disorder Detection
Home
Prediction1
Prediction2
Performance
Genome Disorder Detection


Genome Disorder Detection
Home
Prediction1
Prediction2
Performance


Genome Disorder Detection
Home
Prediction1
Prediction2
Performance





9 Future Enhancement

Further genetic disorders and more prediction models can be added to this study in the future.

10 Conclusion

Technological progress in artificial intelligence has had a significant effect on biological study. In this research, we applied the machine learning model to the original AGDPM model. Information on genetic anomalies was gathered from an online source, and the XGBoost model was used to develop the AGDPM. The model's efficacy was measured using a wide variety of statistical criteria. AGDPM has a higher prediction accuracy than ResNet-50 (92.65%) for identifying diseases caused by mutations in a single gene, mitochondrial disorders, and multifactorial diseases. By aiding in the forecasting of genetic abnormalities, the AGDPM will propel biomedical study forward by leaps and bounds. Additional forecasts and genetic abnormalities may be added to this study in the future to generate a more precise prediction model.

11 References

1. Mc Kusick-Nathans Institute of Genetic Medicine.Online Mendelian Inheritance in Man Johns Hopkins University School of Medicine. Accessed: Nov.1, 2021.Available: www.ncbinlmnih.gov/omim.
2. B. Irom, ‘‘Genetic disorders: A literature review,’’ Genet. Mol. Biol. Res., vol. 4, no. 2, p. 30, 2020.
3. A.Krizhevsky, I. Sutskever, and G. E.Hinton, ‘‘ImageNet classification with deep convolutional neural networks,’’ Commun. ACM, vol. 60, no. 2, pp. 84–90, Jun. 2012.
4. S. J. Sanders, ‘‘First glimpses of the neurobiology of autism spectrum dis-order,’’ Current Opinion Genet. Develop. vol. 33, pp. 80–92, Aug. 2015.
5. Europe PMC Funders Group, ‘‘Biological insights from 108 schizophrenia-associated genetic loci,’’ Nature, vol. 511, no. 7510, pp. 421–427, Jul. 2014.
6. J.Menche, A. Sharma, M. Kitsak, S. D. Ghiassian, M. Vidal, J. Loscalzo, and A.-L. Barabasi, ‘‘Uncovering disease-disease relationships through the incomplete interactome,’’ Science, vol. 347, no. 6224, Feb. 2015, Art. no. 1257601.
7. A. L. Barabási, N. Gulbahce, and J. Loscalzo, ‘‘Network medicine: A network-based approach to human disease,’’ Nature Rev. Genet., vol. 12, pp. 56–68, Oct. 2011.
8. M. Vidal, M. E. Cusick, and A.L. Barabási,‘‘Interactome networks and human disease,’’ Cell, vol. 144, no. 6, pp. 986–998, Mar. 2011.
9. X.Wang, N.Gulbahce, and H. Yu, ‘‘Network-based methods for human disease gene prediction,’’ Briefings Funct.Genomics, vol. 10, no. 5, pp. 280 293, 2011.
10. T.-P. Nguyen and T.B. Ho, ‘‘Detecting disease genes based on semi-supervised learning and protein–protein interaction networks,’’ Artif. Intell. Med., vol. 54, no. 1, pp. 63–71, Jan. 2012.
11. P. Yang, X. L. Li, J. P. Mei, C. K. Kwoh, and S. K. Ng, ‘‘Positive-unlabeled learning for disease gene identification,’’ Bioinformatics, vol. 28, no. 20, pp. 2640–2647, 2012.
12. A. Rishabh. Of Genomes and Genetics HackerEarth Machine Learning Challenge. Kaggle. Accessed:Oct. 27, 2021. Available: https://www.kaggle.com/aryarishabh/of-genomes-and-genetics-hackerearth-ml-challenge.
13. P. Han, P. Yang, P. Zhao, S. Shang, Y. Liu, J. Zhou, X. Gao, and P. Kalnis, ‘‘GCN-MF: Disease-gene association identification by graph convolutional networks and matrix factorization,’’ in Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Jul. 2019, pp. 705–713.
14. X. Zeng, Y. Liao, Y. Liu, and Q. Zou,‘‘Prediction and validation of disease genes using HeteSim scores,’’ IEEE/ACM Trans. Comput. Biol. Bioinf., vol. 14, no. 3, pp. 687–695, May 2017.
15. H. Zhou and J. Skolnick,‘‘A knowledge-based approach for predicting gene–disease associations,’’ Bioinformatics, vol. 32, no. 18, pp. 2831–2838, Sep. 2016.
16. Y. Li, H. Kuwahara, P. Yang, L. Song, and X. Gao, ‘‘PGCN: Disease gene prioritization by disease and gene embedding through graph convolutional neural networks,’’ bioRxiv, vol. 2019, Jan. 2019, Art. no. 532226, doi: 10.1101/532226.
17. K. Yang, Y. Zheng, K. Lu, K. Chang, N. Wang, Z. Shu, J. Yu, B. Liu, Z. Gao, and X. Zhou, ‘‘PDGNet: Predicting disease genes using a deep neural net-work with multi-view features,’’ IEEE/ACM Trans. Comput. Biol. Bioinf., vol. 19, no. 1, pp. 575–584, Jan. 2022, doi: 10.1109/TCBB.2020.3002771.
18. M. Alshahrani and R. Hoehndorf, ‘‘Semantic disease gene embeddings (SmuDGE): Phenotype-based disease gene prioritization without pheno-types,’’ Bioinformatics, vol. 34, no. 17, pp. i901–i907, Sep. 2018.
19. K. Yang, R. Wang, G. Liu, Z. Shu, N. Wang, R. Zhang, J. Yu, J. Chen, X. Li, and X. Zhou, ‘‘HerGePred: Heterogeneous network embedding represen-tation for disease gene prediction,’’ IEEE J. Biomed. Health Informat., vol. 23, no. 4, pp. 1805–1815, Jul. 2019.

20. K. Yang, N. Wang, G. Liu, R. Wang, J. Yu, R. Zhang, J. Chen, and X. Zhou, ‘‘Heterogeneous network embedding for identifying symptom candidate genes,’’ J. Amer. Med. Inform. Assoc., vol. 25 Nov. 2018.
21. Y. Liu, H.Q. Qu, X. Chang, L. Tian, J. Qu, J. Glessner, P. M. A. Sleiman, and H. Hakonarson, ‘‘Machine learning reduced gene/non-coding RNA features that classify schizophrenia patients accurately and highlight insightful gene clusters,’’ Int. J. Mol. Sci., vol. 22, no. 7, p. 3364, Mar. 2021.
22. Y. Liu, H.Q. Qu, F. D. Mentch, J. Qu, X. Chang, K. Nguyen, L. Tian, J.Glessner, P. M. A. Sleiman, and H. Hakonarson, ‘‘Application of deep learning algorithm on whole genome sequencing data uncovers structural variants associated with multiple mental disorders in African American patients,’’ Mol. Psychiatry, vol. 27, no. 3, pp. 1469–1478, Mar. 2022, doi: 10.1038/s41380-021-01418-1.
23. Rectifier/ (Neural/ Networks). Accessed:Nov. 4, 2021.
24. Statistics#03—Standard Deviation and Variance. Accessed: Nov. 4, 2021.Available: https://towardsdatascience.com/statistics-03-standard-deviation-and-variance-9724f33b58df.
25. Softmax Activation Function—How It Actually Works. Accessed: Nov.4, 2021. Available: https://towardsdatascience.com/softmax-activation-function-how-it-actually-works-d292d335bd78
26. A.-U. Rahman, S. Abbas, M. Gollapalli, R. Ahmed, S. Aftab, M. Ahmad, M.A. Khan, and A. Mosavi, ‘‘Rainfall prediction system using machine learning fusion for smart cities,’’ Sensors, vol. 22, no. 9, p. 3504, May 2022.
27. M. Saleem, S. Abbas, T. M. Ghazal, M. A. Khan, N. Sahawneh, and M.Ahmad, ‘‘Smart cities: Fusion-based intelligent traffic congestion con-trol system for vehicular networks using machine learning techniques,’’ Egyptian Informat. J., vol. 6, pp. 1–10, Apr. 2022.
28. M. W. Nadeem, H. G. Goh, M. A. Khan, M. Hussain, M. F. Mushtaq, and V. A. Ponnusamy,‘‘Fusion-based machine learning architecture for heart disease prediction, ’’ Comput. Master. Continue vol. 67, no. 2, pp. 2481–2496, 2021.
29. S. Y. Siddiqui, A. Athar, M. A. Khan, S. Abbas, Y. Saeed, M. F. Khan, and M. Hussain, ‘‘Modelling, simulation and optimization of diagnosis cardiovascular disease using computational intelligence approaches,’’J.Med. Imag. Health Informat., vol. 10, no. 5, pp. 1005–1022, May 2020.
30. N. Taleb, S. Mehmood, M. Zubair, I. Naseer, B. Mago, and M. U. Nasir, ‘‘Ovary cancer diagnosing empowered with machine learning,’’ in Proc. Int. Conf. Bus. Anal. Technol. Secur. (ICBATS), Feb. 2022, pp. 1–6.
31. A.-U. Rahman, A. Alqahtani, N. Aldhafferi, M. U. Nasir, M. F. Khan, M.A. Khan, and A. Mosavi, ‘‘Histopathologic oral cancer prediction using oral squamous cell carcinoma biopsy empowered with transfer learning,’’ Sensors, vol. 22, no. 10, p. 3833, May 2022.