International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 11 Issue: 11 | Nov 2024 www.irjet.net p-ISSN: 2395-0072

A DEEP LEARNING APPROACH TO DETECT GENETIC BASED DISEASE IN PREGNANCY PERIOD

Elavarasi T[1], Mariappan P[2]

[1]*PG Student, Department of Computer Science and Engineering, Velalar College of Engineering and Technology, Erode, Tamil Nadu, India*
[2]*Assistant Professor, Department of Computer Science and Engineering, Velalar College of Engineering and Technology, Erode, Tamil Nadu, India*

Abstract - *The identification of genetic disorders during early pregnancy remains a challenging task in medical research, despite the use of Deep Learning techniques. Current methods for identification have accuracy issues, which are further complicated by the lack of labelled and unlabelled data. To tackle this issue, semi-supervised approaches like label propagation and positive-unlabelled learning are used to identify candidate disease genes, utilizing unknown genes for training. Recent advancements in Deep Learning and diagnostic imaging have revolutionized computerized healthcare, opening up new possibilities in disease identification during pregnancy. To enable early detection of genetic diseases, an Improved Algorithm and statistically significant text information are incorporated. Key clinical information such as age, sex, and genes are included, while resting state functional data is used to measure brain connectivity. Deep Learning techniques are employed for data interpretation and analysis, facilitating the classification of variations and data models. This approach offers better solutions for identifying genetic disorders during pregnancy. The Convolutional Neural Network algorithm is particularly effective for checking data in a more compact way with training and testing data, delivering promising results. This promising development in the field of healthcare can lead to early diagnosis and prevention of genetic diseases during pregnancy.*

Key Words: Genetic disease, Convolutional Neural Networks, health care, genetic, pregnancy.

1. INTRODUCTION

The identification of genetic disorder in pregnancy associations is of immense importance in the diagnosis and treatment of human genetic disorder in pregnancy. However, the number of genetic disorder in pregnancy related genes that have been identified and reported in public databases such as the web based and the Genetic Association Database is limited. Therefore, the search for genetic disorder in pregnancy is still crucial. Traditional gene mapping techniques involve linkage analysis and comprehensive association studies. However, due to the limited number of crossovers in tested families, linkage analysis only identifies chromosomal regions that may contain dozens or even hundreds of candidate genes. Comprehensive association studies may also identify multiple regions that still need to be investigated in future research, making experimental validations of many candidate genes time consuming and expensive.

Since integrating multiple sources of data is essential for identifying genetic disorder in pregnancy related genes, a series of network based computational approaches have been proposed over the past decade. The common idea behind these methods is that genes causing similar or related genetic disorder in pregnancy will be closely related to each other in biological networks. These models typically use text mining of biomedical literature, functional annotations, pathways and ontologies, co expression relationships, intrinsic gene properties, protein interactions, regulatory information, orthologous relationships, and gene expression data to identify candidate genetic disorder in pregnancy [1].

For example, the text mining approaches to identify a large number of human gene networks contained in the database. The genetic disorder in pregnancy associations by using a global network distance measure called random walk analysis will define similarities in protein-protein interaction networks. However, the main limitation of these network based methods is that they fall short of summarizing complex genetic disorder in pregnancy, for which there are no gene linkage studies yet. The potential of Deep Learning lies in the belief that we can replicate the functioning of the human brain by creating the right connections using silicon and wires, much like living neurons and dendrites [2].

The human brain is a complex network of 80 billion nerve cells known as neurons that are connected to thousands of other cells through Axons. These neurons receive inputs from the surrounding environment or physical organs through dendrites, which generate electrical signals that quickly travel through the brain network. Based on these inputs, a neuron can either transmit the message to another neuron to address the issue or choose not to send it.

To address this issue, another approach called Inductive Grid Completion would be applied, which is based on multiple biological sources and can be implemented to genetic disorder in pregnancy is not observed at training

time. Of all the methods for identifying genes relevant to a given genetic disorder in pregnancy, the traditional Integrated Network based Deep Learning approach will perform the best in identification of genetic disorder in pregnancy.

Genetic diagnoses for one family member may have implications for the health of relatives, even without any current symptoms. Precision genomics based medicine has emerged as a means of providing personalised and effective treatment based on patients genetic characteristics. Researchers are aiming to capitalise on advances in genomics to develop increasingly accurate illness risk prediction models to realise the full potential of precision medicine.

The model, derived from a predictor and two classifiers, predicts the presence of genetic disorders and specifies the disorder and disorder subclass, if present. Despite recent progress in polygenic risk scores, the outcomes of these scores are still limited due to present methodologies.

The primary motivation of Genetic disorder in pregnancy identification analysis with Deep Learning on Neural Networks is to detect the genetic disorder in pregnancy in the hospital dataset. In this work, the dataset containing the patient dataset will be taken into consideration. The pre processing will be applied in to the dataset and the noisy and null value data will be removed from the dataset. After the data will be analysed and visualized for further processing. The Convolutional Neural Networks algorithm will be chosen to implementation process. The project evaluation can be tested with the Deep Learning algorithm prediction results. Since the Convolutional Neural Networks algorithm will be used to predict the genetic disorder in pregnancy, the accuracy of the algorithm result will be helpful to evaluate the results. The accuracy score of the algorithm in the Genetic disorder in pregnancy identification helps to evaluate the dataset.

The Deep Learning will be the python based application which contributes to find out the Genetic disorder in pregnancy early stage. It will be helpful for the human to detect at early and to take necessary treatments in the correct time. The progression of profound learning influences is generally applied to classify the assignments and portrayals learning. These profound frameworks with numerous layers have been displayed to yield promising execution in removing serious areas of strength for more of information. The streamlining of the goal capability becomes curved in the event that we adjust one variable and fix the others.

The finding of the application includes the 'Clinical Elements' and 'Clinical Administration' segments of the website pages that report the side effects, prescription and reactions by patients, and related investigations of impacts of various courses of treatments.

Since cross approval on review information presumably prompts overoptimistic results, cross approval is improper for this issue. To assess the capacity of the models to anticipate new found affiliations.

Thus, via consistently consolidating the model for helper side data and the cooperative filter for the quality sickness affiliations grid, the model learns more significant portrayal for every quality and illness and gives more exact expectation. The Deep Learning algorithm will be used to predict the genetic disorder in pregnancy, the accuracy of the algorithm result will be helpful to evaluate the results [3]. The accuracy score of the algorithm in the Genetic disorder in pregnancy identification helps to evaluate the dataset.

1.1 Objective

- The objective of Genetic disorder in pregnancy identification with Deep Learning is to detect the Genetic disorder in pregnancy in the early stage itself with the available attributes.
- In this work, the hospital patient dataset will be taken into consideration.
- The pre processing will be applied in to the dataset and the noisy and null value data will be removed from the dataset.
- After the data will be analysed and visualized for further processing.
- The Deep Learning neural network algorithm will be chosen to make the good accuracy prediction.
- The aspect of correlation coefficient data is less sensitive to genetic disorder in pregnancy compared to the genetic dataset.

Distinct attributes of genetic disorder in pregnancy occurrence can be removed on varying scales to achieve greater accuracy in performance.

1.2 Related Work

A highlight determination strategy was implemented to decrease the number of atomic descriptors in a fair and unbiased manner. The strategy involved two stages, namely statistical analysis and Genetic Algorithm. In the first stage, descriptors with low standard deviation or containing similar values over half were removed. Subsequently, a Pearson correlation analysis was carried out to determine the relationship among the descriptors and between the descriptor and the target. This step was performed to reduce bias and eliminate descriptors with redundant information. Descriptors with weak correlation with the target (correlation < 0.1) or strong correlation with another objective (correlation > 0.9) were removed [4].

In case of overlapping descriptors, the one with a weaker correlation with the target was eliminated. In the second stage, a combination of descriptors was chosen using the Genetic Algorithm technique. This technique follows Darwin's principles of natural evolution and uses random methods to obtain optimal non random solutions. The descriptor selection by GA was performed by defining the solution as a collection of a whole number value in a chromosome. In this case, the number of the value is equal to the number of the selected descriptor, where the value represents the descriptor list. The cross entropy loss was used as a performance metric during the feature selection [5].

The prediction model uses the Genetic Algorithm technique, which resembles the structure and function of the natural neural system. The primary principle of the Genetic Algorithm is the implementation of artificial neurons, which are simple mathematical models. Such a model has three simple sets of rules, namely reproduction, mutation, and activation. The y-scrambling analysis ensures that the performance of the model did not correlate with an accidental relationship. This analysis was conducted by shuffling the class centre while preserving the descriptors multiple times. The results of the y-scrambling demonstrate by providing the values for shuffled and unshuffled data [6] [7].

2. PROPOSED SYSTEM

The proposed methods aim to find the genetic disorder in pregnancy with higher standard. The accuracy levels of the identification of the genetic disorder in pregnancy will be improved with the proposed system. The Deep Learning on neural network will provide the better solution to solve the problem of identification of the genetic disorder in pregnancy in the real world hospital data. The Convolutional Neural Network algorithm will check the data in more compact with training and testing the data. It will provide more accuracy as compared with the other type of techniques. The genetic patient dataset will be taken as the input to the application and the dataset will be passed into the Convolutional Neural Network algorithm and the data will be analysed with the different visual graphs.

The proposed approach applies a biased neural network function with effectiveness, enabling reliable recognition of genetic disorder in pregnancy. Figure 1 shows the proposed architecture diagram, the Genetic disorder in pregnancy dataset was given as input to the application and the pre processing was applied, next the data cleaning was performed after the training and test data were split down and will be passed into Deep Learning algorithm and the Genetic disorder in pregnancy will be predicted. The Genetic disorder in pregnancy identification with Deep Learning on neural networks will be the python based application which contributes to find out the genetic disorder in pregnancy. It will be helpful in finding of the genetic disorder in pregnancy based on the attributes of the patient records.

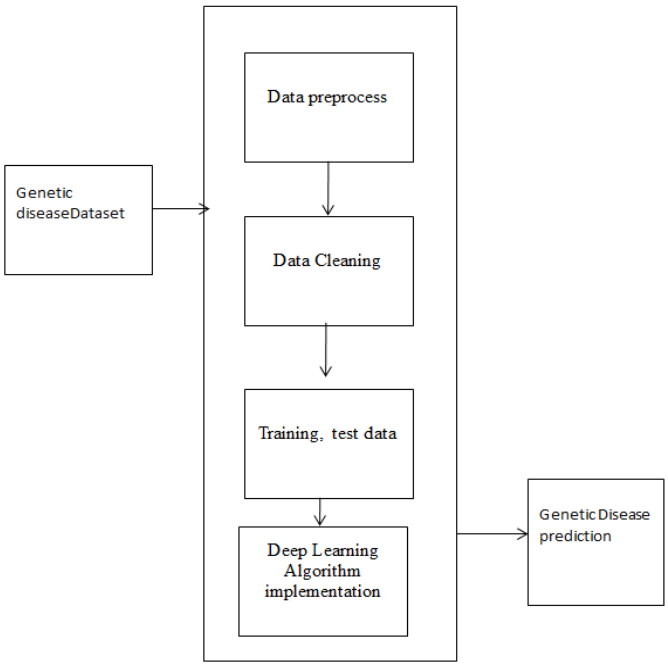

Fig -1: Proposed Architecture diagram

The testing and training variables were split and passed into the algorithm for the Genetic disorder in pregnancy prediction. In this algorithm will provide a comprehensive and intelligent solution for discovering high utility item sets, enabling users to access important information and streamline the search processes.

The application will be developed with Google Colab Python Tool as the project can be directly executed in any type computer systems with internet connection. There was no need of any specific software to be installed in the user system. The Colab Tool helps to develop and run the application directly inside the cloud server where the Python library files were installed. The Deep Learning algorithm libraries were built inside the Colab.

2.1 Data Description

The dataset for genetic disorder in pregnancy identification was taken from the source of Kaggle dataset. This dataset contain the fields needed for the analysing of the patient dataset. Exploratory examination is a cycle to investigate and comprehend the information and information relationship in a total profundity with the goal that it makes highlight designing and Deep Learning demonstrating steps smooth and smoothed out for expectation. Exploratory examination assists with validating the presumptions or misleading. Most of the image in a dataset were noisy and contain lots of information. But with feature engineering do, will get more good results. The first step was to import the

libraries and load data. After that will take a basic understanding of data like its shape, sample, were there any NULL values present in the dataset. Understanding the data is an important step for prediction or any Deep Learning project. It is good that there were no NULL values. The Figure 2 shows the Sample Training Dataset.

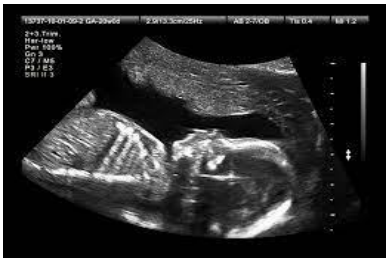

Fig -2: **Sample Training Dataset**

The information about the genetic patient records with different types of attributes were collected from Kaggle data. The dataset total contains of image dataset with training and testing genetic affected images of patients. It will begin from the principal segment and investigate every section and comprehend what influence it makes on the objective segment. The dataset of brain scan images are downloaded from Kaggle website. It has 3 folders train, test, Val which has genetic affected, non affected patient brain x-ray images. Figure 3 shows the Sample Testing Dataset

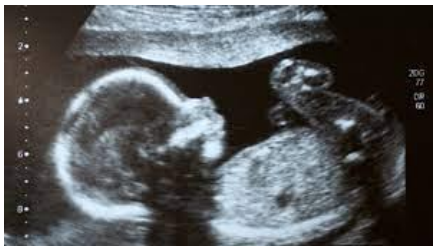

Fig -3: **Sample Testing Dataset**

At the necessary step, we will likewise perform pre processing and include designing undertakings. The point in acting top to bottom exploratory examination is to get ready and clean information for better Deep Learning demonstrating to accomplish elite execution and summed up models. So it should begin with breaking down and setting up the dataset for expectation.

3. EXPERIMENTAL ANALYSIS

The initial process of loading the dataset into the Google Colab into the drive is the first step in execution process. The image data containing the information of the image with respect to the path and the description of the image location and the image related style were linked. The pre processing was applied to the dataset where all the noisy data were removed and the image was reshaped as per the mapping of 255 pixel.

The information has an extremely straightforward design with elements. Each folder is related with the Genetic disorder in pregnancy scan images. The image was reshaped with the following the protocol of making the size of the image to 255 pixel range in any format types. The noisy data present inside the image was also removed and improves the image quality which will be more helpful in the application of the prediction of the genetic disorder in pregnancy in the dataset.

The image dataset is divided into testing and training to pass in to the neural network model.

The Keras Models Programming interface provides a versatile platform for constructing intricate neural networks by adding and removing layers. The Application Programming Interface (API) supports both sequential and functional models with a single input and output or multiple inputs and outputs, respectively. The training module encompasses various methods, including generating the model, optimizer, and loss function, fitting the model and evaluating and predicting input data. Furthermore, the API includes methods for batch data processing, testing, and prediction. The models programming interface in Keras also enables users to save and pre process the models for future use. Therefore, this API offers an efficient and comprehensive solution for building and training neural networks. The Keras library files were applied into the execution process.

Heat maps use colour to display data magnitude, enabling easy identification of patterns and anomalies. Brighter, reddish colors are used to represent more common or higher activity values, while darker colours were used to represent less common values or activity. The shading matrix used to define the heat map is also commonly referred to as the heat map itself. The seaborn.heatmap () function were used to plot heat maps.

The heat map matches the genetic disorder in pregnancy is present (yes) and not present (no) values in the given dataset. The attributes of input data was been trained and tested then the model have been built, trained and tested. The results of the Convolutional Neural Network (CNN) with sequential data analysis provide the accuracy. The validation and testing accuracy were identified while executing the CNN algorithm.

The CNN algorithm was applied with creating the sequential model. The results of the prediction of genetic disorder in pregnancy identification with the CNN provide the accuracy results. The training and validation of the model were evaluated and the accuracy was calculated.

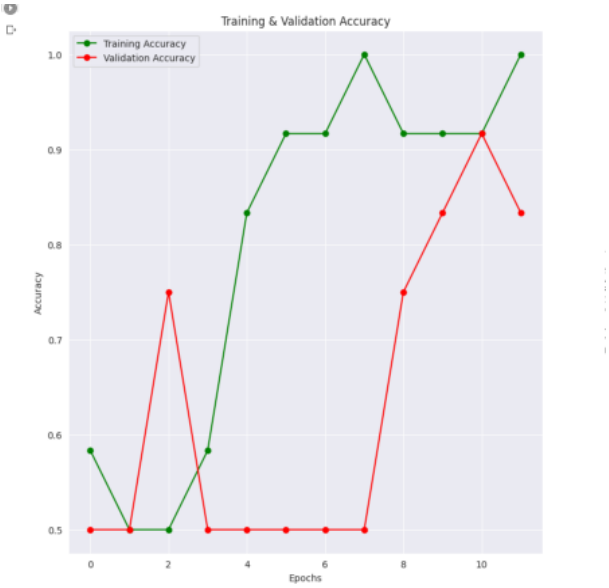

Fig - 4: Training and Validation accuracy

The Figure 4 shows the training and validation accuracy results in the graphical format. The training accuracy is getting in the increase ratio and it reaches the good saturation point. The validation accuracy is getting in the gradual increase points and reaches the average of 85% of accuracy.

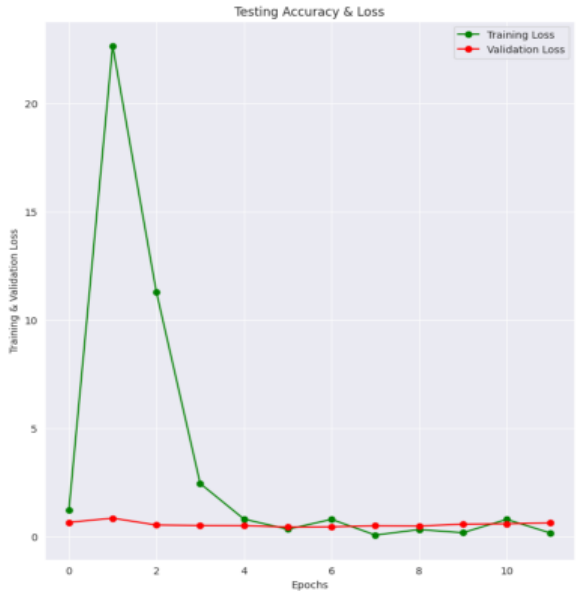

Fig - 5: Testing and accuracy loss

The training loss is getting in the decrease ratio and it reaches the good saturation point as shown in Figure 5. Thus the genetic disorder in pregnancy identification accuracy is calculated to make the prediction quality good.

4. CONCLUSIONS

A cutting edge framework for detecting genetic disorder in pregnancy has been developed using Deep Neural Networks and diverse medical data. The framework employs all Ultra-sound scan images with genetic information for model training and data classification. By constructing functional intellectual networks based on signal correlation, the neural network formation is optimized using correlation coefficient information. This methodology greatly enhances diagnostic accuracy compared to traditional approaches, demonstrating that integrating Advanced Deep Learning with medical expertise is an effective way to diagnose neurological disorders in their early stages. The same or similar methodologies can be applied to diagnose other neurological genetic disorder in pregnancy, providing a foundation for ongoing diagnosis in this field. Assessing the effectiveness of Deep Learning techniques and algorithms for forecasting genetic disorders and their sub categories can enhance the model's precision. Additionally, scrutinizing the dataset has enabled to identify the most suitable feature sets for model fitting. Health departments, clinics, and hospitals can utilize the model for real medical diagnosis, while the study findings may prove valuable in genetic disorder lab experiments. Enhancing accessibility and usability can be achieved through the addition of a Graphical Interface, such as website applications. The application of genetic disorder in pregnancy predictive models in varied clinical populations can enhance the performance and limitations of the proposed models, thus refining medical practice. Though prediction remains a challenge, future research is promising and may provide a wealth of clinically useful information if evaluated within the appropriate context.

REFERENCES

[1] Jiajie Tang, Jin Han, Jiaxin Xue, Li Zhen, Xin Yang, Min Pan, Lianting Hu, Ru Li, Yuxuan Jiang, Yongling Zhang, "A Deep-Learning-Based Method Can Detect Both Common and Rare Genetic Disorders in Fetal Ultrasound". Biomedicines 2023, 11, 1756. https://doi.org/10.3390/biomedicines11061756.

[2] Gonzaludo, N., Belmont, J.W., Gainullin, V.G., Taft, R.J.," Estimating the burden and economic impact of pediatric genetic disease". Genet. Med. 2019, 21, 1781–1789.

[3] Arnaout, R., Curran, L., Zhao, Y.; Levine, J.C., Chinn, E., Moon-Grady, A.J. "An ensemble of neural networks provides expert-level prenatal detection of complex congenital heart genetic disorder in pregnancy". Nat. Med. 2021, 27, 882–891.

[4] Basel-Vanagaite, L, Wolf, L, Orin, M, Larizza, L, Gervasini, C, Krantz, I.D, Deardoff, M.A, "Recognition of the Cornelia de Lange syndrome phenotype with facial

dysmorphology novel analysis". Clin. Genet. 2016, 89, 557–563.

[5] Lumaka, A., Cosemans, N., Mampasi, A.L., Mubungu, G., Mvuama, N., Lubala, T., Mbuyi-Musanzayi, S.,Breckpot, J., Holvoet, M., De Ravel, T. "Facial dysmorphism is influenced by ethnic background of the patient and of the evaluator". Clin. Genet. 2017, 92, 166–171.

[6] Miyagi, Y., Hata, T., Bouno, S., Koyanagi, A., Miyake, T.," Recognition of facial expression of fetuses by artificial intelligence (AI)". J. Perinat. Med. 2021, 49, 596–603.

[7] Valentine, M.; Bihm, D.C.; Wolf, L.; Hoyme, H.E.; May, P.A.; Buckley, D.; Kalberg, W., Abdul-Rahman, O.A. "Computer-aided recognition of facial attributes for fetal alcohol spectrum disorders". Pediatrics 2017, 140, e20162028.

[8] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D. Grad- CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In Proceedings of the IEEE International Conference on Computer Vision IEEE 2017, Venice, Italy, 22–29 October 2017; pp. 618–626.

[9] Frid-Adar, M.; Diamant, I.; Klang, E.; Amitai, M.; Goldberger, J.; Greenspan, H. GAN-based Synthetic Medical Image Augmentation for increased CNN Performance in Liver Lesion Classification. Neuro computing 2018, 321, 321–331.

[10] J. T. Eppig, J. A. Blake, C. J. Bult, J. A. Kadin, and J. E. Richardson, "The Mouse Genome Database (MGD): new features facilitating a model system". Nucleic Acids Research, vol. 35, issue, pp. 630–7, 2007.

[11] S. S. Dwight, M. A. Harris, K. Dolinski, C. A. Ball, G. Binkley, K. R. Christie, D. G. Fisk, L. Issel-Tarver, M. Schroeder, and G. Sherlock, "Saccharomyces genome database (SGD) provides secondary gene annotation using the Gene Ontology (GO)," Nucleic Acids Research, vol. 30, no. 1, pp. 69–72, 2002.

[12] T. L. Saito, M. Ohtani, H. Sawai, F. Sano, A. Saka, D. Watanabe, M. Yukawa, Y. Ohya, and S. Morishita, "SCMD: Saccharomyces cerevisiae morphological database".NucleicAcidsResearch,vol.32, no. 1, pp. 319–22, 2004.

[13] K. L. Mcgary, I. Lee, and E. M. Marcotte, "Broad network-based predictability of saccharomyces cerevisiae gene loss-of-function phenotypes". Genome Biology, vol. 8, no. 12, p. R258, 2007.