Indonesian Journal of Electrical Engineering and Computer Science
Vol. 33, No. 1, January 2024, pp. 227~235
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v33.i1.pp227-235

A novel two-tier feature selection model for Alzheimer's disease prediction

Sonam V. Maju, Gnana Prakasi Oliver Sirya Pushpam[2]
Department of Computer Science and Engineering, Christ (Deemed to be University), Bangalore, India

Article Info

Article history:

Received Jun 30, 2023
Revised Sep 25, 2023
Accepted Oct 13, 2023

Keywords:

Hybrid regression models
Machine learning algorithms
neural network
Regression model
Statistical validation

ABSTRACT

The interdisciplinary research studies of artificial intelligence in health sector is bringing drastic life saving changes in the healthcare domain. One such aspect is the early disease prediction using machine learning and regression algorithms. The purpose of this research is to improve the prediction accuracy of Alzheimer 's disease by analysing the correlation of unexplored Alzheimer causing diseases. The work proposes Chi square-lasso ridge linear (Chi-LRL) model, a new two-tier feature ranking model which recognizes the significance of including diabetes, blood pressure and body mass index as potential Alzhiemer predictive parameters. The newly added predictive parameters of Alzheimer's disease were statistically verified along with the conventional prediction parameters using chi-square method (Chi) as Tier 1 and an embedded model of lasso, ridge and linear (LRL) Regression for feature ranking as Tier 2. The performance of the proposed Chi-LRL model with selected features were then analysed using machine learning algorithms for performance analysis. The result shows a noticeable performance by selecting eleven significant features and a 4.5% increase in the prediction accuracy of Alzheirmer disease.

Corresponding Author:

Sonam V. Maju
Department of Computer Science and Engineering, Christ (Deemed to be University)
Bangalore, India
Email: sonam.maju@res.christuniversity.in

1. INTRODUCTION

The traditional perspectives of decision making have been reformed with the development of prediction algorithms. Present algorithms leverage the analysis methods and escalate the performance by acting as a catalyst to the prediction process in every sector. In healthcare, the intervention of artificial intelligence (AI) is actually saving millions of lives [1]. The prediction models of AI can be widely used for diseases such as Alzheimer's disease (AD), Bi-Polar disorder and Autism, which cannot be cured but can be treated if predicted early. Alzheimer's disease affects the daily routine of the patients [2] and the affected person even struggles to complete the day-to-day activities. An effective treatment in curing AD is early prediction the main research gap in predicting Alzheimer disease is to identify the right predictive paramaters and the multi collinearity problem among predictive parameters during feature selection process. This research attempts to predict the occurrence of AD by understanding its correlation with other health conditions. A possible way for this is to consider the diseases that trigger the development of AD. Unfortunately, the existence of diseases or symptoms which trigger the condition of AD are available in the electronic health records (EHR) but have not been scientifically explored completely yet. Different case studies prove the correlation of high blood sugar and AD [3]–[5]. Hence, as a new parameter, blood sugar level is considered as one of the parameters that may contribute to an individual being diagnosed with AD in the future. The diabetic parameter that was used in this

research work is Glyhb (A1C test) and is available in the EHR [6]. Therefore, diabetes can be used as one parameter because any individual with diabetes and A1C test value greater than 10 is highly diabetic and can trigger the condition of AD in due course of time [7]. Other new parameters that trigger the condition of Alzheirmer's include blood pressure [8]–[10], body mass index (BMI) [11], [12] and cholesterol [13], [14]. From the studies, this research uses diabetes, blood pressure, cholesterol and body mass index as the additional predictive parameters of Alzheiemer disease along with the conventional parameters. Before using these parameters into prediction directly, a feature ranking of these parameters should be done and understand the significance of each parameter. Later, these parameters can be selected based on the ranking and significance for the disease prediction. Thus, the objectives of this research study are to identify the new prediction parameters that will help in the early prediction of Alzheimer's, to establish the significance of newly added parameters through feature ranking and to analyse the performance of the proposed two-tier model in predicting Alzheimer's with existing methods.

The data used in this paper is the open access series of imaging studies (OASIS) longitudinal data [15]. As the dataset contains only the conventional parameters for predicting Alzheimers, the new parameters were incorporated from Biostatistics program at Vanderbilt [16]. Hence the dataset contains, age, gender, education, socio economic status, mini-mental state examination (MMSE) score (range is from 0 [worst] to 30 [best]), clinical dementia rating (CDR). (0 = no dementia, 0.5 = very mild AD, 1 = mild AD, 2. = moderate AD), atlas scaling factor (ASF), normalized whole brain volume (nWBV). Whereas, along with the above-mentioned parameters, the second dataset contains, Glyhb/A1C test (Diabetic test), body mass index (BMI) values, and cholesterol. The process in this research began with the identification of additional parameters, the statistical validation of the process is to prove the significance of the newely added parameters and then an ensemble regression model for feature ranking to show how important the newely added parameters. The detailed explanation of using Chi-square method and ensemble reression model is explained in the following sections.

2. METHOD

The few main challenges of clinical research are, the Extraction and identification of prediction parmeters of a disease [17], Multicollinearity between the selected parameters [18], feature ranking and prediction accuracy. The proposed Chi-LRL model, tried to address the above-mentioned challenges throughout the process of Alzheimer prediction. The identification of significant risk factors of Alzheimer's disease was done using a statistical model and processed the Multocollinearity problem with an ensemble regression model. Thus, a two-tier feature ranking model is proposed in this research work. The architecture of two-tier approach is shown in Figure 1 and the two Tiers are described as,

- Tier 1: Chi-square method to identify the significant features.
- Tier 2: An ensemble regression approach for feature selection and ranking.

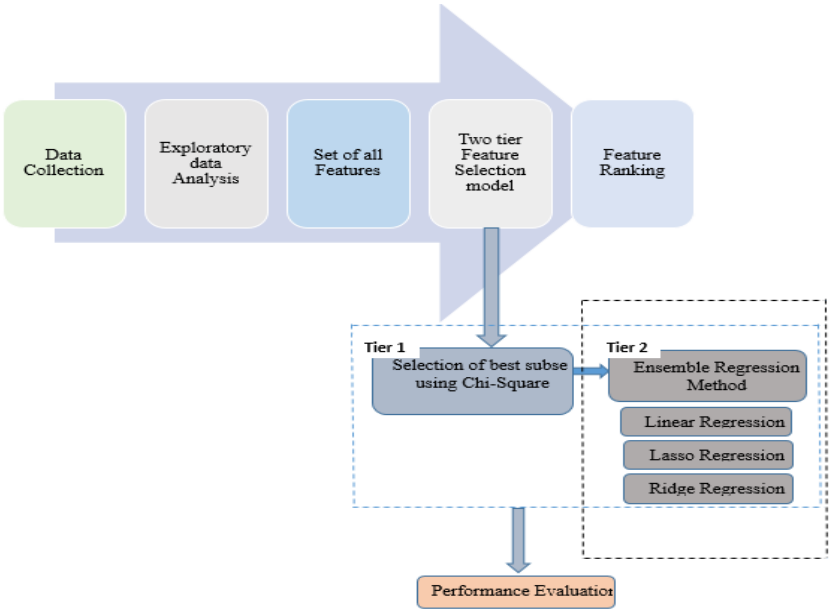

Figure 1. Architecture of Chi-LRL two – tier approach

Initially, a complete set of 25 Alzheimer prediction parameters were used which leads to the problem of dimensionality. In order to avoid that, the most significant features from the total of 25 parameters were selected using a two-tier feature ranking model. Tier 1: Chi-Square method will select the subset of features and Tier 2: an ensemble regression model is used to identify the most significant prediction parameters and rank the parameters accordingly. The chi-square method extracts the first few significant parameters and multicollinearity between these parameters are calculated. Then more features are extracted using the regression model and the feature rankings are calculated to identify the most significant clinical parameter of Alzheimer disease. The Algorithm 1 of the proposed model is shown:

Algorithm 1. Pseudocode of Chi- LRL two-tier model

Input: D: Dataset
F: Set of all features
N: Size of Population
P: Chi-Square Value of each features
n: Number of regression models
χ: Score of each feature using Linear Regression
β: Score of each feature using rank of Ridge Regression
δ: Score of each feature using Lasso Regression
Output: Subset of significant features
    Step 1: Read the dataset and extract all the features into F
     for i=1 to F
     Calculate Chi-square value of each feature
using the following formula and store it in list P.
     $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$
     end for
    Step 2: Load the independent Feature Set F
    Calculate VIF of each Feature to get the multicollinearity values.
$VIF_i = \frac{1}{1-R_i^2}$

    Create a list R to store the selected features
    Fit the regression algorithms and train the model
     for 1 to F
     $F_n = \chi_n + \beta_n + \delta_n$
     end for
    Step 3: Calculate the mean ranks
     for i=1 to F
     $\frac{\chi+\beta+\delta}{3}$
    Load the mean scores into R and sort it.
    Displaying the ranks of each feature in the sorted format based on Ranks.
    End

2.1. Tier 1: Chi- square approach

The approach starts by calculating the significance of each available parameter using Chi square test. The Chi-Square Test assesses associations between categorical variables. Here, a chi-square method is used to calculate the values of 'P', which explains the significance of each feature corresponds to the target (Alzheimer) variable.

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Where, χ^2: Chi squared, O_i: Observed value, E_i: Observed value

- If the P value: 0.05<P<0.10, it is of suggestive significance.
- If the P value: 0.01<P ≤ 0.05, its moderately significant.
- If P value: P≤0.01, its strongly significant. The observed value of P is shown in Table 1.

Table 1. Statistical analysis of clinical parameters

Variables	Diabetic demented	Non-diabetic non-demented	Total	P value
CDR	0.59±0.02	0.01±0.00	0.29±0.02	<0.001**
eTIV	1482.9±12.41	1494.32±13.41	1488.68±9.14	0.533
nWBV	0.72±0.00	0.74±0.00	0.73±0.00	<0.001**
ASF	1.20±0.01	1.19±0.01	1.20±0.01	0.645
Cholesterol	205.82±3.17	211.11±3.36	208.49±2.31	0.253
Stab.glu	113.56±4.56	105.14±3.60	109.29±2.90	0.147
HDL (mg/dl)	51.58±1.30	49.57±1.26	50.56±0.90	0.266
Cholesterol/HDL ratio	4.37±0.11	4.67±0.14	4.52±0.09	0.101
Glyhb	9.45±0.14	4.77±0.07	7.08±0.15	<0.001**

From the Table 1, it is clearly evident that Tier 1 identified only four significant features including the newly added Glyhb parameter along with CDR, MMSE and nWBV as the P value is <0.001. Alzheimer's cannot be predicted accurately with only these four parameters. Hence the work needs to identify additional significant parameters and rank them. So, in the Tier 2, the selection of more significant parameters and ranking is done through an ensemble regression model, which ranks and eliminates the less prominent features.

2.2. Tier 2: ensemble (LRL) regression method

Regression analysis is one statistical method used in understanding the relationship between variables and can evaluate the importance of each independent parameter on the dependent parameter (Alzheimer) [19]. When single regression models are used for feature ranking in Clinical application, because of the correlation of clinical parameters and regularization behaviour, the performance of feature ranking was inaccurate. Hence, single regression models are not compatible for feature selection in our Prediction model. So, an ensemble regression model is used by incorporating the pros of each individual regression models.

As our clinical parameters are highly correlated, the variance inflation factor (VIF) is calculated to check the multicollinearity among the parameters using,

$$VIF_i = \frac{1}{1-R_i^2}$$

VIF can be calculated by dividing one with the tolerance. The tolerance can be calculated by taking subtracting the R squared value from 1. The majority of the clinical parameters used in this research is showing multicollinearity as shown in Figure 2. From the figure, the multicollinearity among the features such as nWBV, ASF, MMSE can be observed with the higher VIF values. If there is multicollinearity between independent features in a multiple regression model, it will be difficult to understand the individual impact of each variable on the target variable.

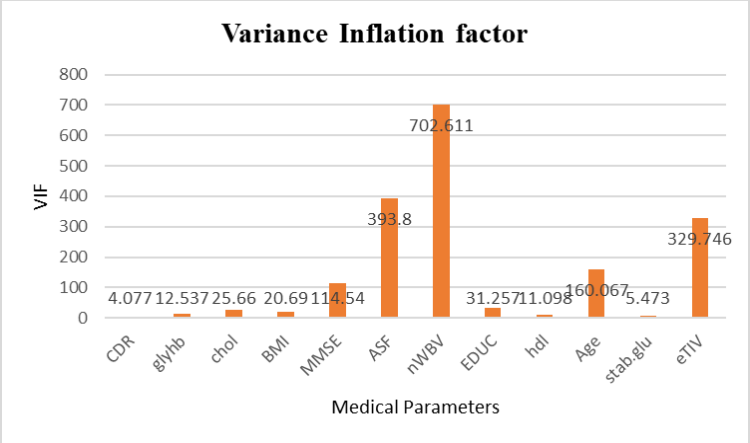

Figure 2. Variance inflation factor

So, in order to solve the problem of multi collinearity, ridge and lasso regression are used by reducing standard errors and shrinking the data to a certain point. In linear regression, prediction error is decomposed based on biased and variation, whereas in ridge, multicollinearity is handled through Shrinkage. Ridge is incapable of feature selection whereas lasso gives the option of feature selection. Ridge will shrink the value coefficients but will never reach zero but lasso can shrink the coefficients to zero. So in case of multicollinearity, the lasso will pick only one attribute thereby shrinking all other correlated attributes to zero. So the proposed ensemble model is the combination of linear regression, lasso and ridge will be used for ranking the features from the dataset. As the behaviour of the dataset is linear, linear regression with recursive elimination factor were used, which recursively do the feature selection process until all the significant features got selected. Later the shrinkage to zero limitation of lasso will be addressed by combining lasso with Ridge and the observed weights will be calculated. The weighted ranks of each feature will be taken and the mean of the weights are calculated accordingly.

The proposed two-tier regression model successfully addresses the problem of dimensionality reduction, multi-collinearity and clinical feature ranking by effectively selecting all the prominent Alzhiemer predictive parameters. Now, these selected features are used in the random forest prediction model for the

performance analysis to prove that these parameters are highly significant and can be used as the prediction parameters of Alzheimer disease. The results and the comparative study of the two-tier feature selection model and random forest prediction model is discussed in the next section.

3. RESULTS AND DISCUSSION

Hence, through the proposed Chi-LRL two tier model, eleven parameters out of 25 parameters were selected whereas the chi square method were able to fetch only four significant features. Later, for selecting all the potential Alzheimer predictive parameters, an ensemble regression model, 'LRL' were used. Inorder to understand the effectiveness of the ensemble model, a performance analysis study on individual regression models were done and is plotted in Figures 3-5. The results of the feature ranking with single regression model states that most of the potential predictive parameters were ranked zero, implicating the least significance towards the target variable. Only 3-4 features will be selected in each Regression models individually. Precisely, from Figure 2, linear regression will extract only four features Figure 3 shows that lasso regression is selecting only two parameters and similarly Figure 4 shows that ridge is selecting only four parameters.

As single regression models were not effective enough to fetch all predictive parameters, an ensemble regression model is used for feature ranking and all the eleven significant predictive parameters of Alzheimer's. The results obtained can be seen in Figure 6 and the features extracted were, M/F, Age, MMSE, CDR, eTIV, nWBV, ASF, cholesterol, HDL, glyhb and BMI. Compared to single regression models represented in Figures 2-4, when an ensemble method is used, every feature will be ranked effectively and these ranks can be used as weights for further disease prediction models. A comparative analysis was done to understand the performance of the ensemble approach in the conventional dataset the result obtained is shown in Figure 6. By analysing Figures 6 and 7, it is very clear and transparent that the newly added features like glyhb, cholesterol and BMI is also ranked and weighted along with other conventional parameters.

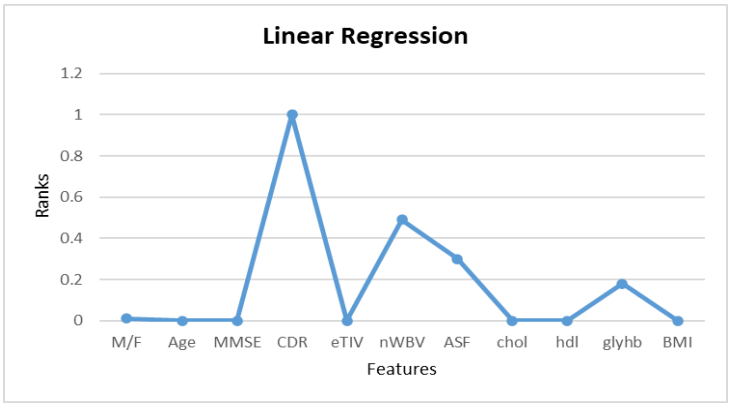

Figure 3. Feature ranking through linear regression

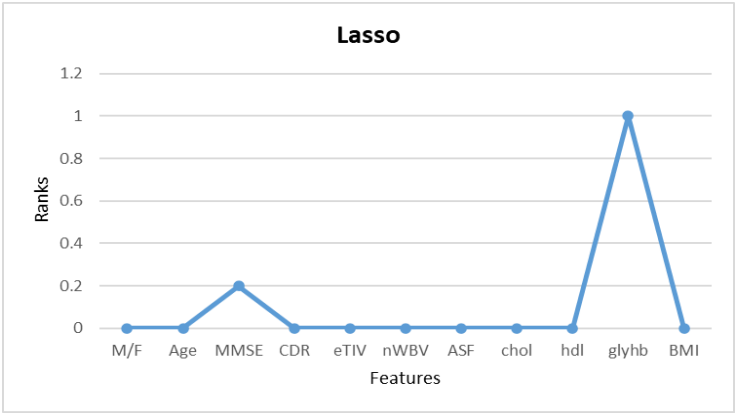

Figure 4. Feature ranking through lasso regression

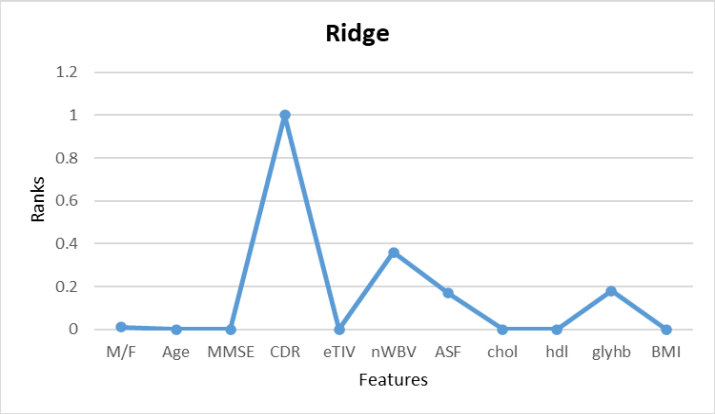

Figure 5. Feature ranking through ridge regression

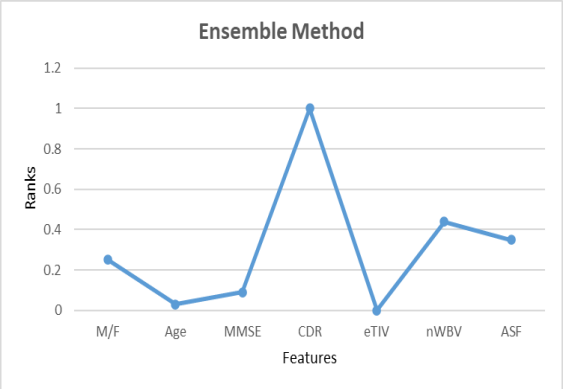

Figure 6. Feature ranking with hybrid regression analysis on conventional dataset

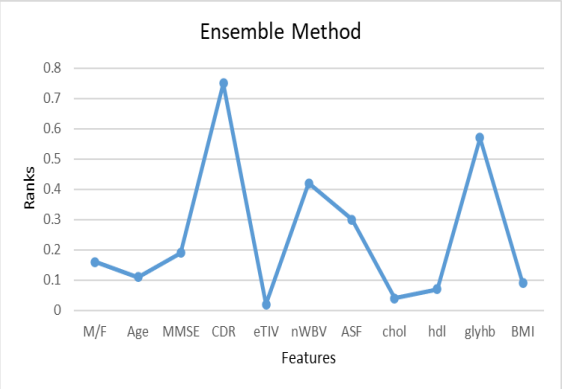

Figure 7. Feature ranking with hybrid regression analysis on dataset with additional parameters

Table 2 gives the comparative analysis on different feature selection methods and the number of features extracted. The dataset used in all these feature selection methods are the OASIS dataset with 15 features and almost all the methods extracted common 5 features, MMSE, CDR. nWBV, eTIV and ASF. Also, our Chi-LRL two tier selection model outperforms majority of the existing methods by identifying a greater number of significant features like glyhb, hdl and cholesterol which had not been identified and utilized in the existing models. Hence, our model gave an adaptable performance by extracting 11 features through a two-tier model by addressing the complexities of multicollinearity. Later, a comparative study is being done using random forest classifier for the Alzheimer prediction on the conventional parameters and with the conventional and newly added parameter. The performance metrics observed were accuracy, recall and AUC score.

Table 2. Comparative analysis on feature extraction

Reference	Feature selection method	Features extracted
An improved multi-modal based machine learning approach for the prognosis of Alzheimer's disease [20]	Chi-squared statistical test, feature importance and correlation matrix	M/F, Age, EDUC, SES, MMSE, eTIV, nWBV, and ASF
Prediction of Alzheimer's disease (AD) using machine learning techniques with boruta algorithm as feature selection method [21]	Boruta algorithm with random forest grid search cross validation	MMSE, CDR, eTIV, nWBV and ASF
Early-Stage Alzheimer's disease prediction using machine learning models [22]	Correlation coefficient, Information gain, and Chi-Square	M/F, Age, EDUC, SES, MMSE, eTIV, nWBV, ASF, CDR
Dementia prediction on OASIS dataset using supervised and ensemble learning techniques [23]	Heat map and multi-collinearity	Gender, Age, EDUC, MMSE, CDR. nWBV, eTIV

From the observed results, it is clearly evident that the accuracy of the dataset with additional parameters outperforms the conventional dataset by 3%. The addition of new parameters directly helps in the better prediction of Alzheimer disease. The results obtained are shown in Figure 8. So, the accuracy performance is undoubtedly increased when more predictive parameters were added for the Alzheimer prediction. So, a comparative study has been done on Machine learning as well as deep learning models. Table 3 and Figure 9, shows the comparative analysis of the prediction models on two different datasets, OASIS and ADNI. The highest accuracy observed in OASIS dataset is 92% and received a 6% increase in the accuracy when used new Prediction parameters in the same dataset. The ADNI dataset gave 93% accuracy with Convolutional neural networks. Hence, can say the addition of new prediction parameters and usage of right significant features can increase the performance of every prediction model.

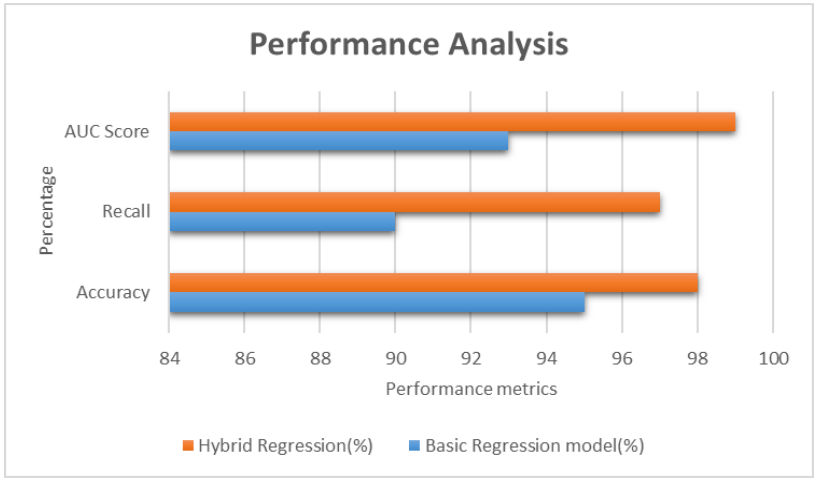

Figure 8. Performance metrics

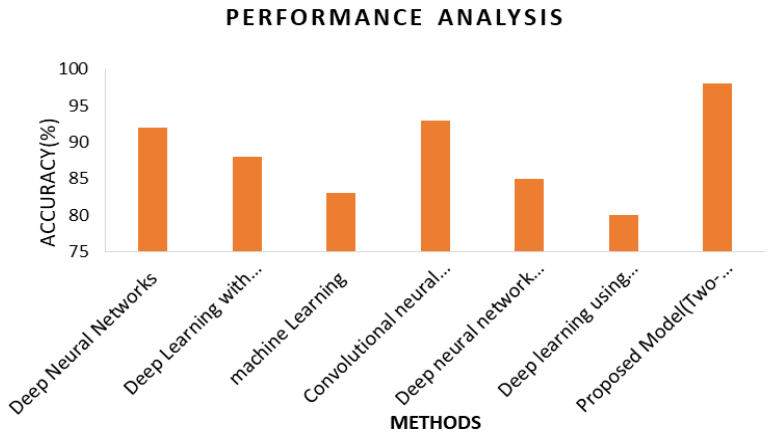

Figure 9. Performance analysis with other existing methods

Table 3. Comparative analysis of accuracy metric

Reference	Dataset	Methods	Accuracy
Kavitha *et.al.* [22]	OASIS dataset	Machine learning models	83%
Vinayak *et.al.* [23]	OASIS dataset	Machine learning on MRI images	97.8
Basheer *et. al.* [24]	OASIS dataset	Deep neural networks	92%
Saratxaga *et al.* [25]	OASIS dataset	Deep learning and image processing technique	88%
Helaly *et al.* [26]	ADNI dataset	Convolutional neural networks	93%
Prajapati *et al.* [27]	ADNI dataset	Deep neural network binary classifier	85%
Martinez *et al.* [28]	ADNI dataset	Deep learning using convolutional auto encoders	80%
Sudharsan and Thailambal [29]	ADNI dataset	Machine learning models	75%
Maju and Prakasi [30]	OASIS dataset	Machine learning models	98%

While comparing the proposed Chi-LRL model with the existing models, the proposed model outperforms mainly in three research gaps, identification of a new prediction parameter and proving its significance through our feature selection model, considering the multi collinearity aspect in the feature selection and the increase in the accuracy of Alzheimer prediction with the addition of new parameters and usage of the right features. Hence, the performance of the proposed model is addressing solutions for the existing research gaps in feature engineering and disease prediction.

4. CONCLUSION

The proposed model identified new predictive parameters like, cholesterol, HDL, glyhb and BMI from the EHR for the better prediction of Alzheimer disease. In addition, the proposed ensemble model outperformed the single regression models by ranking the following parameters M/F, age, MMSE, CDR, eTIV, nWBV, ASF, cholesterol, HDL, glyhb and BMI as significant predictive parameters of Alzheimer disease. Thus, the performance of the chi-LRL model increased by 4.5% in prediction accuracy, as self explanatory for the utilization of adding more features for disease prediction. The graphical visualizations given in this research paper gives a clarity of using ensemble and hybrid models over Single individual models. Hence, the proposed Chi-LRL model, addresses the above-mentioned challenges throughout the process of Alzheimer prediction. When other related papers fail to incorporate newly added parameters. Thus, the proposed Chi-LRL model successfully used the parameters and observed better accuracy. The scope of the research work can be improved by identifying more significant features directly from electronic health records using named entity recognition of natural language processing for better decision making.

REFERENCES

[1] A. Ramesh, C. K. Subbaraya, and R. K. G. Krishnegowda, "A remote health monitoring framework for heart disease and diabetes prediction using advanced artificial intelligence model," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 30, no. 2, pp. 846–859, 2023, doi: 10.11591/ijeecs.v30.i2.pp846-859.

[2] M. W. Bondi, E. C. Edmonds, and D. P. Salmon, "Alzheimer's disease: Past, present, and future," *Journal of the International Neuropsychological Society*, vol. 23, no. 9-10 Special Issue, pp. 818–831, Oct. 2017, doi: 10.1017/S135561771700100X.

[3] T. T. Nguyen, Q. T. H. Ta, T. K. O. Nguyen, T. T. D. Nguyen, and V. Van Giau, "Type 3 diabetes and its role implications in alzheimer's disease," *International Journal of Molecular Sciences*, vol. 21, no. 9, p. 3165, Apr. 2020, doi: 10.3390/ijms21093165.

[4] M. Michailidis, D. Moraitou, D. A. Tata, K. Kalinderi, T. Papamitsou, and V. Papaliagkas, "Alzheimer's Disease as Type 3 Diabetes: Common Pathophysiological Mechanisms between Alzheimer's Disease and Type 2 Diabetes," *International Journal of Molecular Sciences*, vol. 23, no. 5, p. 2687, Feb. 2022, doi: 10.3390/ijms23052687.

[5] O. Nisar, H. Pervez, B. Mandalia, M. Waqas, and H. K. Sra, "Type 3 diabetes mellitus: A link between alzheimer's disease and type 2 diabetes mellitus," *Cureus*, Nov. 2020, doi: 10.7759/cureus.11703.

[6] S. Chatterjee and A. Mudher, "Alzheimer's disease and type 2 diabetes: A critical assessment of the shared pathological traits," *Frontiers in Neuroscience*, vol. 12, no. JUN, Jun. 2018, doi: 10.3389/fnins.2018.00383.

[7] H. J. Lee, H. I. Seo, H. Y. Cha, Y. J. Yang, S. H. Kwon, and S. J. Yang, "Diabetes and Alzheimer's disease: mechanisms and nutritional aspects," *Clinical Nutrition Research*, vol. 7, no. 4, p. 229, 2018, doi: 10.7762/cnr.2018.7.4.229.

[8] C. J. Lee *et al.*, "lBlood pressure levels and risks of dementia: A nationwide study of 4.5 million people," *Hypertension*, vol. 79, no. 1, pp. 218–229, Jan. 2022, doi: 10.1161/HYPERTENSIONAHA.121.17283.

[9] R. A. A. de Heus *et al.*, "Association between blood pressure variability with dementia and cognitive impairment: A systematic review and meta-analysis," *Hypertension*, vol. 78, no. 5, pp. 1478–1489, Nov. 2021, doi: 10.1161/HYPERTENSIONAHA.121.17797.

[10] M. H. Cho, *et al.* "Blood pressure and dementia risk by physical frailty in the elderly: a nationwide cohort study". *Alzheimers Research & Therapy*, vol.15, March 2023. doi: 10.1186/s13195-023-01211.

[11] J. A. Flores-Cordero, A. Pérez-Pérez, C. Jiménez-Cortegana, G. Alba, A. Flores-Barragán, and V. Sánchez-Margalet, "Obesity as a risk factor for dementia and Alzheimer's disease: the role of leptin," *International Journal of Molecular Sciences*, vol. 23, no. 9, p. 5202, May 2022, doi: 10.3390/ijms23095202.

[12] S. Charisis *et al.*, "Obesity impacts the expression of Alzheimer's disease–related genes: The framingham heart study," *Alzheimer's and Dementia*, vol. 19, no. 8, pp. 3496–3505, Aug. 2023, doi: 10.1002/alz.12954.

[13] F. M. Feringa and R. van der Kant, "Cholesterol and Alzheimer's disease; from risk genes to pathological effects," *Frontiers in Aging Neuroscience*, vol. 13, Jun. 2021, doi: 10.3389/fnagi.2021.690372.

[14] V. R. Varma *et al.*, "Abnormal brain cholesterol homeostasis in Alzheimer's disease—a targeted metabolomic and transcriptomic study," *npj Aging and Mechanisms of Disease*, vol. 7, no. 1, p. 11, Jun. 2021, doi: 10.1038/s41514-021-00064-9.

[15] D. S. Marcus, A. F. Fotenos, J. G. Csernansky, J. C. Morris, and R. L. Buckner, "Open access series of imaging studies: Longitudinal MRI data in nondemented and demented older adults," *Journal of Cognitive Neuroscience*, vol. 22, no. 12, pp. 2677–2684, Dec. 2010, doi: 10.1162/jocn.2009.21407.

[16] S. Gill and P. Pathwar, "Prediction of diabetes using various feature selection and machine learning paradigms". *Modern Approaches in Machine Learning & Cognitive Science: A Walkthrough*. *Studies in Computational Intelligence*, vol. 1027.pp 133-146.April 2022. doi: 10.1007/978-3-030-96634-8_12.

[17] M. Z. I. Chowdhury and T. C. Turin, "Variable selection strategies and its importance in clinical prediction modelling," *Family Medicine and Community Health*, vol. 8, no. 1, p. e000262, Feb. 2020, doi: 10.1136/fmch-2019-000262.

[18] K. I. Sundus, B. H. Hammo, M. B. Al-Zoubi, and A. Al-Omari, "Solving the multicollinearity problem to improve the stability of machine learning algorithms applied to a fully annotated breast cancer dataset," *Informatics in Medicine Unlocked*, vol. 33, p. 101088, 2022, doi: 10.1016/j.imu.2022.101088.

[19] K. M. Adeniyi, O. J. Oladapo, T. O. Araoye, T. F. Adebayo, S. V. Egoigwe, and M. C. Odo, "Prediction of heart disease outcomes using machine learning classifier," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 30, no. 2, pp. 917–926, 2023, doi: 10.11591/ijeecs.v30.i2.pp917-926.
[20] A. Khan and S. Zubair, "An improved multi-modal based machine learning approach for the prognosis of Alzheimer's disease," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 6, pp. 2688–2706, Jun. 2022, doi: 10.1016/j.jksuci.2020.04.004.
[21] L. K. Leong and A. A. Abdullah, "Prediction of Alzheimer's disease (AD) using machine learning techniques with boruta algorithm as feature selection method," *Journal of Physics: Conference Series*, vol. 1372, no. 1, p. 012065, Nov. 2019, doi: 10.1088/1742-6596/1372/1/012065.
[22] C. Kavitha, V. Mani, S. R. Srividhya, O. I. Khalaf, and C. A. T. Romero, "Early-stage Alzheimer's disease prediction using machine learning models," *Frontiers in Public Health*, vol. 10, Mar. 2022, doi: 10.3389/fpubh.2022.853294.
[23] S. S. Vinayak E, A. Shahina, and A. K. Nayeemulla, "Dementia prediction on OASIS dataset using supervised and ensemble learning techniques," *International Journal of Engineering and Advanced Technology*, vol. 10, no. 1, pp. 244–254, Oct. 2020, doi: 10.35940/ijeat.a1827.1010120.
[24] S. Basheer, S. Bhatia, and S. B. Sakri, "Computational modeling of dementia prediction using deep neural network: Analysis on OASIS dataset," *IEEE Access*, vol. 9, pp. 42449–42462, 2021, doi: 10.1109/ACCESS.2021.3066213.
[25] C. L. Saratxaga *et al.*, "Mri deep learning-based solution for alzheimer's disease prediction," *Journal of Personalized Medicine*, vol. 11, no. 9, p. 902, Sep. 2021, doi: 10.3390/jpm11090902.
[26] H. A. Helaly, M. Badawy, and A. Y. Haikal, "Deep learning approach for early detection of Alzheimer's disease," *Cognitive Computation*, vol. 14, no. 5, pp. 1711–1727, Sep. 2022, doi: 10.1007/s12559-021-09946-2.
[27] R. Prajapati, U. Khatri, and G. R. Kwon, "An efficient deep neural network binary classifier for Alzheimer's disease classification," in *3rd International Conference on Artificial Intelligence in Information and Communication, ICAIIC 2021*, Apr. 2021, pp. 231–234, doi: 10.1109/ICAIIC51459.2021.9415212.
[28] F. J. Martinez-Murcia, A. Ortiz, J. M. Gorriz, J. Ramirez, and D. Castillo-Barnes, "Studying the manifold structure of Alzheimer's disease: A deep learning approach using convolutional autoencoders," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 1, pp. 17–26, Jan. 2020, doi: 10.1109/JBHI.2019.2914970.
[29] M. Sudharsan and G. Thailambal, "Alzheimer's disease prediction using machine learning techniques and principal component analysis (PCA)," *Materials Today: Proceedings*, vol. 81, pp. 182–190, 2021, doi: 10.1016/j.matpr.2021.03.061.
[30] S. V. Maju and O. S. G. Prakasi, "Utilization of dark data from electronic health records for the early detection of Alzheimer's disease," in *Recent Advances in Artificial Intelligence and Data Engineering. Advances in Intelligent Systems and Computing*, Springer, 2022, pp. 195–203.

BIOGRAPHIES OF AUTHORS

Sonam V. Maju received the B.tech. degree in Information technology engineering from Mahatma Gandhi University, Kerala, India in 2013 and the M.Tech. degree in Computer Science and Engineering from Mahatma Gandhi University, Kerala, India in 2015. She is currently pursuing the Ph.D. degree in Computer science and engineering at Christ University, Bangalore, India. Her research interest includes the machine learning, natural language processing, information management, and clinical research. She had published one Indian patent and several other publications on Clinical Research. She can be contacted at email: sonam.maju@res.christuniversity.in.

Dr. Gnana Prakasi Oliver Sirya Pushpam Assistant professor at the CHRIST (Deemed to be University), Department of Computer Science and Engineering, Bangalore, India. She received her B.Tech. degree in Information Technology and M.E. degree in Software Engineering from Anna University. Her research interests include mobile Ad hoc networks, IoT, software engineering and machine learning. Major research focus in MAC and Network layers and published research papers in reputed journals. Currently focusing on analyzing the prediction algorithms using recurrent neural network with clinical data. She has more than 15 publications in refereed national and international journals. She can be contacted at email: gnana.prakasi@christuniversity.in.