A FRAMEWORK FOR PERFORMANCE EVALUATION OF MACHINE LEARNING TECHNIQUES TO PREDICT THE DECISION TO CHOOSE PALLIATIVE CARE IN ADVANCED STAGES OF ALZHEIMER'S DISEASE

Mutyala Sridevi
Research Scholar, VTU Research Centre and Asst. Professor, Department of MCA,
BMS Institute of Technology & Management, Bangalore, India
sridevim@bmsit.in
https://bmsit.irins.org/profile/116012

Arun Kumar B.R.
Professor, Department of Computer Science and Engineering and Research Supervisor, VTU Research Centre, Department of MCA,
BMS Institute of Technology & Management, Bangalore, India
arunkumarbr@bmsit.in
https://bmsit.irins.org/profile/116014

Abstract - Alzheimer's is one of the chronic diseases that stand as a challenge in the geriatrics domain. The symptoms, diagnosis and treatment varying from person to person makes it more complex to understand the inherent nature of this disease. The lifestyle factors and behavioural traits play major role in the onset and progression of Alzheimer's disease compared to genetic factors. Studying and analysing such behavioural traits would help the healthcare practitioners to understand its effect on the progression of the disease and its manageability in the advanced stages. This would culminate in enhancing the provision of customized palliative care to alleviate the trauma faced by patients and their caregivers. Making the machines learn from and build models on such data would make the task of healthcare professionals, much easier and quicker. It assists the healthcare stakeholders in studying large amounts of patient data and get equipped to work with patient specific symptoms, to strategize their treatment and improve terminal care facilities. The social media data, available in the form of discussions on patient support groups hosted by Facebook, is manually curated and used for training various machine learning models. Gaussian Naïve Bayes followed by Support Vector Machine were found to be the best performing models based on various context specific evaluation metrics.

Keywords: Social media; Alzheimer's; behavioural traits; machine learning; evaluation metrics; feature selection.

1. Introduction

Alzheimer's is an irreversible and a progressive brain disorder that gradually hampers the memory and thinking skills which culminates into cognitive impairment [1]. This disease is more prevalent in western countries compared to other parts of the world. Despite being ranked as one of the successful and happiest countries in the world, Finland and United States stand at the top among the most affected countries, by Alzheimer's. Canada, Iceland, and Sweden follow in numbers raising debatable questions on the factors that lead to this disease. Alzheimer's has been the 6th leading cause of death in the United States though the mortality rates have significantly improved in the last 20 years [2]. Magnetic Resonance Imaging (MRI) features that depict the brain tissue shrinkage is the only means, so far, to predict the rate at which the cognitive health declines from mild impairment to dementia. But, researchers believe that the interplay of lifestyle, environmental and behavioural factors constitute the major cause for onset and progression of Alzheimer's, which eventually affect the brain, though less than 1% of the time, it is caused by gene specific factors [3]. Studying and using the lifestyle related factors, personality & behavioural traits, and their effect on the important aspects of Alzheimer's disease will be of great help to the medical practitioners (specifically neurologists and neuropsychologists) to customize their treatment and counselling methods that alleviate the trauma this disease causes to the patients as well as caregivers. This kind of data would even help them to understand various trends related to the disease and its progression.

Patient networks/communities on social media are a rich source of data in the form of discussions about how patients and their caregivers deal with the disease, its progression and repercussions, diagnosis and treatment methods, and many such relevant issues. Analysing this data would help the stakeholders in this domain to make an in depth study about this disease and strategize their treatment methods as the onset, symptoms, and progression of Alzheimer's varies from patient to patient and is highly dependent on their lifestyle, relationship with the caregiver, and other factors. In this setting, our current work focuses on analysing the behavioural traits in association with some demographic factors of the Alzheimer's patients that are collected from the discussions of their caregivers on social media patient networks. This work is a step towards finding the best performing machine learning model to predict whether the caregivers opt to take care of their loved ones at home i.e., whether the disease is manageable in the advanced stages or the caregivers resort to palliative care/hospice based on the patient behaviour and other relevant factors. This would provide an informative outlook about the disease to the healthcare stakeholders as to how various attributes play significant role in the manageability of the patients in their last stages.

1.1 *Relevant work*

Vast amount of work was being carried out in applying machine learning techniques in the healthcare domain to predict various diseases. The authors in [4] have made an extensive literature review on the studies made to compare performances of various supervised machine learning algorithms that are used in disease prediction. Nearly 336 unique articles were published in indexed journals that discussed about the application of machine learning in prediction of diseases like breast cancer, heart disease, Parkinson's etc. The work in [5] presented a systematic review of the tools for data analysis in the healthcare domain by providing examples of various machine learning algorithms. The authors of [6] reviewed briefly, the important literature on how deep learning can benefit the researchers to diagnose the Alzheimer's disease at early stages. As the diagnosis of Alzheimer's disease mostly uses brain images that visualize the pharmacology, functionality, and structure of the brain, which are generated by using imaging techniques like functional and structural imaging [7], all the Alzheimer's related research rely on this kind of data only.

A mathematical tool called path signature for the selection of features for machine learning to predict a diagnosis for the Alzheimer's disease was discussed in [8]. Ubiquitous technologies generate time-ordered data through patient monitoring methods from which predictive features can be selected by using the path signature method. This research work used whole brain, hippocampus, and ventricles related measurements as variables and further feature selection. Though some research towards using deep learning techniques in the diagnostic classification and detection of Alzheimer's is being carried out recently, it is very limited and moreover, they mostly rely on neuroimaging data and fluid biomarkers for their analysis [9]. The research study presented in [10] applied a binary classification benchmarking algorithm to predict the risk of late-onset of Alzheimer's disease by using data related to genetic variation. It also carried out systematic comparisons of machine learning algorithms' performance in classification which would help to harness the predictive capabilities of the models discussed in their work. The work by the authors in [11] presented the development of a machine learning model which is used to predict dementia in general patient population from different institutions of healthcare by making use of routine care data that is available as electronic health records. Structured and unstructured data from diagnosis, medical notes, and prescriptions was used to train their model proposed and predict one and three years prior to the onset of the disease. Easily understandable ML techniques were selected by the authors to make it adaptable by the healthcare professionals.

The data related to Alzheimer's disease diagnosis, classification, and prediction that is being analysed so far, is in the form of neuro-images and biomarkers. But, Alzheimer's patients also exhibit some behavioural disturbances and psychological symptoms as prominent features observed in approximately 90% of the patients suffering from it. The symptoms like anxiety, aggression, excessive sleep, and other abnormalities in behaviour can be noticed commonly. The process of neuro-degeneration is affected by dysfunctionalities in neurotransmission along with personality traits and psychological factors playing an equally progressive role [12]. A study indicated only moderate differences between the types of mental or behavioural traits noticed at different phases of illness caused by Alzheimer's that leads to different types of dementia. In [13], large community of elderly people were investigated to find that Alzheimer's patients are more likely to have illusions, deviant motor behaviour, and aggression as common symptoms and less likely to suffer from depression. The behavioural deviations from normal are generally described as "agitation" which include wandering, aggression etc. [14]. Analysing such behaviour related data of the Alzheimer's patients by applying machine learning models is not taken as a research problem up to date.

1.2 *Our contribution*

Analysing data shared on social media related to behavioural traits of Alzheimer's patients has a huge potential of making the stakeholders of healthcare domain stay well-equipped to make important/informed decisions related to diagnosis, strategize treatments and provide more effective palliative care for the victims. The details about the behavioural traits of Alzheimer's patients along with other demographic and related factors would serve as a rich source of data, and when analysed, would unveil some interesting trends in understanding the problem at depth. Our current work focuses on designing a framework to make the machine learn from such data gathered from online patient communities' discussions (caregivers as members), evaluate its performance with various models and their goodness of fit to the data at hand. This would enable the deployment of applications developed based on this framework across institutions for its extensive usage. The data in the form of discussions on social media patient support groups is self-reported by the caregivers who directly deal with the patients and is at par with any proctored survey data in which, actually, the scope of data collection has many limitations. Forthcoming parts of this paper present the proposed conceptual framework in Section 2, data source, collection and experimental setting in Section 3, feature engineering and selection in Section 4, elaboration on the machine learning techniques used in Section 5, results and discussion in Section 6 followed by conclusion and future work.

2. Proposed Conceptual Framework

The steps followed to conduct the experiment and find the best performing model for the given problem and data collected is depicted in Fig. 1. The upcoming sections will further elaborate on each of these steps.

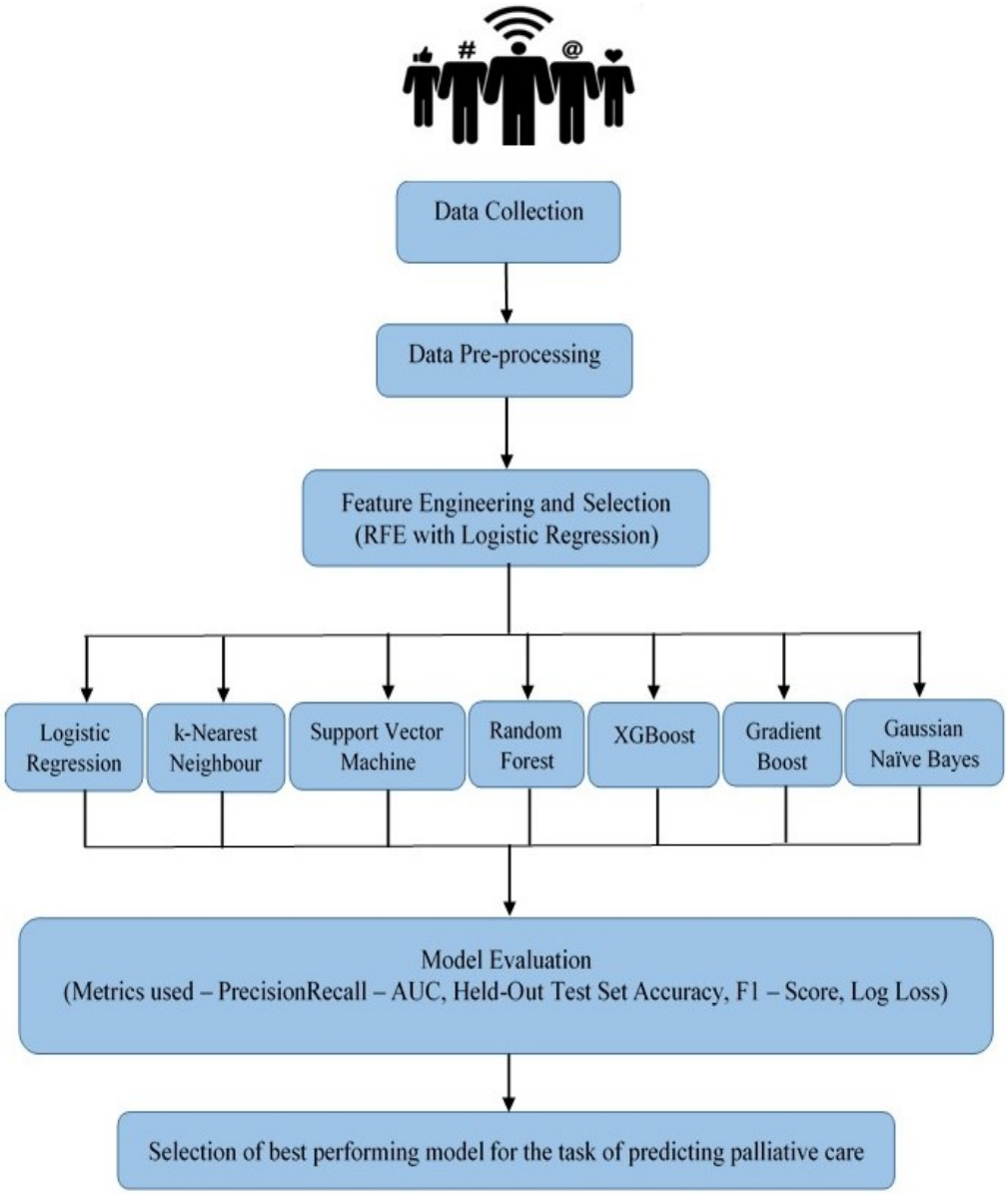

Fig. 1. Framework for the study

3. Data Collection and Experimental Setting

The patient related data and behavioural traits required to train the machine learning models, that can predict the decision of caregivers to opt for a hospice, is collected from the discussions between caregivers on social media patient support networks/communities hosted by Facebook. These communities provide valuable source of such information strengthening the recent news that therapists are proposing online support groups to the patients, for them to obtain emotional support and overcome geographical isolation. Though our general emphasis is on predicting the chronic disease trends by using social media data, the specific focus is directed towards making the machines learn and predict trends in Alzheimer's disease using online caregiver discussions (as patients can't directly participate due to cognitive impairment) because this area didn't receive much attention from the research community. Data like gender of the patient, age at which they are diagnosed with Alzheimer's, country to which he/she belongs to, behaviour of the patient as reported by the caregiver in his/her discussion, in what kind of activity they are engaged in, throughout the day, patient and caregiver biological relation with each other and the frequency of the messages by the caregiver on the patient network in terms of seeking advice or venting out, was collected. The data of 243 patients, manually collected by observing their posts on the timeline, is retained after pre-processing the data to remove missing values and outliers. The patient identification data is anonymised to protect the privacy of the patients as per the data privacy law.

4. Feature Engineering and Selection

The process of data preparation for modelling is coined as "Feature Engineering". Feature engineering is an essential step to achieve good performance for predictive modelling problems and is dependent on how the data is prepared for modelling [15]. The steps include right from exploratory analysis and visualization of the data, handling missing or redundant data, dealing with outliers, encoding categorical features, engineering numeric features, and selecting features that majorly contribute in predicting the outcome/target variable.

The data that is collected is in numerical as well as categorical format. Most of the machine learning models do not work well with categorical data. The Python library "sklearn" used for the machine learning algorithms requires the features to be in numerical format. So, they have to be converted into numerical values with proper encoding techniques suitable to the characteristics of the features. The numerical values are min-max-scaled to bring them into a fixed range, normally 0 to 1. This is done to bring down the standard deviations and suppress the outlier effect. The categorical values are target encoded as obtaining dummies will increase the dimensionality of the dataset with increase in the categories for each feature. Ordinal encoder is another option if there is specific ordering in the categories of the features. The dependent or the predicted variable, that is the decision of choosing a hospice was originally tabulated as Yes or No which is further label encoded to convert it into 1 and 0, respectively. These steps bring all the values of the features into a common range and ready for fitting the machine learning models.

Recursive Feature Elimination (RFE) is one of the feature selection methods that fits a model while removing the less contributing features iteratively until specified number of features is reached [16]. The collinearity and dependencies in the model, if any, will be eliminated by recursively eliminating small number of features in every iteration. Thus features will be ranked by the model. Though the number of features to be retained is not known in advance, 5-fold stratified cross-validation is used along with recursive feature elimination in order to find the optimal number of features. This process scores different subsets of features and selects the best scoring subset of features.

We attempted to use logistic regression (linear model) and random forest (non-linear model), further elaborated in the next section, as estimators for the RFE feature selection algorithm to choose relevant features. In order to automatically get the number of features selected by RFE, cross-validation evaluation of different sets of features is performed and the optimal number of features with the best mean score is selected. The following Fig. 2 shows the optimal number of features with random forest and logistic regression as estimators.

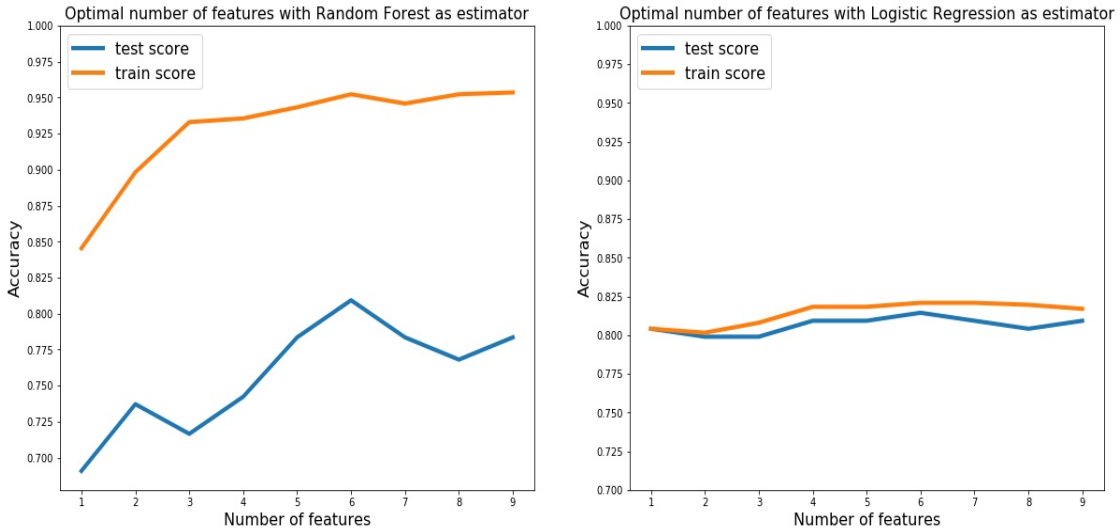

Fig. 2. Optimal number of features using RFE with non-linear and linear estimator

The illustration in Fig. 2 shows that random forest estimator seems to over fit with training score maximizing around 0.95 and test score at 0.81 for 6 features selected after which there is no considerable improvement in the training and test scores. On the other hand, logistic regression's training score is maximum at 0.82 and test score at 0.81 for 6 features, again, with no further improvement. This leads to a decision of selecting logistic regression as estimator for RFE to perform automatic feature selection before choosing various models to fit on the selected features and compare their performance.

Having known that selection of 6 features gives the best 5-fold cross-validated training and test scores, it is now intriguing to know which features are those. The feature ranks as assigned by RFE are presented in Table 1.

Table 1. Ranks assigned by RFE to features (1 – most contributing feature)

Feature description	Rank
Female	1
Male	2
Age Diagnosed	4
Frequency of Messages	1
Behavior	1
Activity Engaged in	1
Care Giver	1
Location	1
Relation	3

5. Machine Learning Techniques used

The selection of machine learning techniques depends on the data source and also the application domain [17]. Models namely Logistic Regression, k-Nearest Neighbour, Support Vector Machine, Random Forest, XG Boost classifier, Gradient Boost classifier, and Gaussian Naïve Bayes were used in the present work in order to find the best performing model by careful fine-tuning of the hyper-parameters. As the data collected is a combination of numerical, categorical as well as binary, it is compelling to understand how each of the above mentioned, proved to be efficient models (in various scenarios) perform with the data concerned.

5.1 *Logistic regression*

Logistic regression is a machine learning algorithm which is used for the classification problems based on the concept of probability to perform predictive analysis and especially when the target variable is of categorical type. It uses a complex cost function defined as 'Sigmoid function' or simply a 'logistic function' to bring predicted values down into the range of 0 and 1 [18]. Sigmoid function is used to map the predicted values to probabilities so that the assignment of data points to the appropriate classes is performed by the algorithm with some amount of likelihood. The equation for Sigmoid function is provided below as Eq. (1).

$$y = 1/(1 + e^{-x}) \quad (1)$$

The Sigmoid function graph is depicted in Fig. 3.

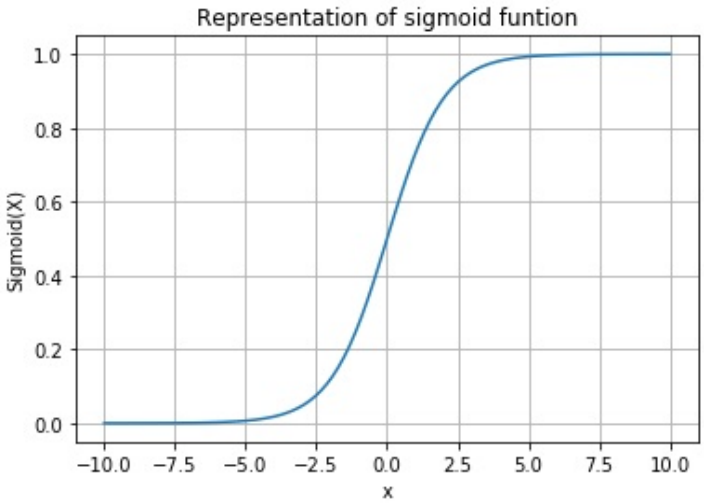

Fig. 3. Illustration of a Sigmoid function

The variable *e* represents the exponential constant with an approximate value of 2.71828. While using the logistic regression, usually a threshold is specified and all values above the set threshold will be assigned to one class and all values below it will be assigned to the other class. There are several solver options available but only 'lbfgs' was selected for the model fitting in this work as it avoids drawbacks like saddle points and is fast compared to other solvers. The maximum number of iterations were chosen as 100 to reach the convergence of the model. Logistic regression can also be used for multiclass classification problems by following the notion of 'one Vs. all'.

5.2 *k-Nearest Neighbour (kNN)*

KNN is a simple supervised machine learning algorithm which can be used to solve both regression and classification problems. This algorithm depends on labelled training data to learn a function which predicts an appropriate output when posed with an unseen/unlabelled data. Depending on the type of problem being solved, a choice among various distance measures like Euclidean, Minkowski etc. can be made to compute the distance between the data points [19]. The hyper parameter *k* has to be carefully tuned to avoid overfitting (model learning too much) or under fitting (model learning too less) of the model and increase its generalization capability.

5.3 *Support Vector Machine (SVM)*

The SVM finds a hyperplane that clearly classifies the sample points in an n-dimensional feature space. Among the many possible hyperplanes that can be chosen, to separate the two classes, a plane that has the maximum margin between the classes will be chosen as the optimal hyperplane. This provides reinforcement to the model so that the unseen data points can be more confidently classified.

The SVM's approach can be seen in Fig. 4. This is good enough for linearly separable classes but to perform complex classification with real data, kernalised SVMs can be used.

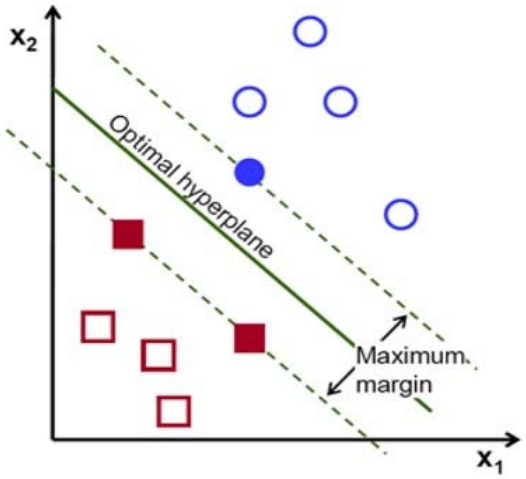

Fig. 4. Approach of Support Vector Machine [20]

Kernalised SVMs take the original input data space and transforms it to a different higher dimensional feature space after which it becomes easier to linearly classify the transformed data. Though there are many transformations available, Radial Basis Function (RBF) kernel proved to be the best performing option and thus selected for the experiment under discussion. The RBF kernel can be represented by the following Eq. (2).

$$K(x, x') = \exp[-\gamma . ||x - x'||^2] \quad (2)$$

A kernel can be defined as similarity measure (dot product) between the data points. The hyper-parameters γ and C have to be tuned properly. Smaller values for γ represent larger similarity radius resulting in decision boundaries that are smooth and larger values of γ require data points to be more closer to be considered as similar. On the other hand, C is a regularization parameter that interacts with the γ parameter. If γ is large, C will not have much influence. If γ is small, the model is heavily constrained and C affects the classifier as it does with a linear classifier. For the present experiment, both γ and C values were set to 1 as it is observed that the classifier gives best performance at these hyper-parameter values.

5.4 *Random forest*

The random forest classifier uses an ensemble of individual decision trees. The individual class predictions by each tree are counted to find the class with maximum votes which will be the model's prediction [21]. The underlying notion can be attributed to 'wisdom of crowds'. The key feature is low correlation between the models in ensemble, as uncorrelated models would produce better ensemble predictions than any of the individual predictions. But, increasing the complexity of the random forest classifier in terms of increasing its number of estimators and maximum depth may result in overfitting of the model that performs too good on the training data and badly on the test set. Careful tuning would help to overcome this problem and reap the benefits of this ensemble method. The number of estimators is set to 5 and maximum depth to 2 to obtain best possible performance from this model.

5.5 *eXtreme Gradient Boost (XGBoost) classifier*

XGBoost is an implementation of gradient boosted decision trees which is available as an open source library. Though basically it is an ensemble technique, this classifier takes an iterative approach. Each new model is trained to predict and correct the mistakes (called residuals) made by the previous models. This process is repeated until it converges i.e., no further improvements in performance can be made. The main objective of this iterative process is to avoid individual models ending up making same kind of mistakes. XGBoost works to optimize the objective function which encourages the predictive models and optimize regularization to achieve lesser variance that makes predictions stable. The objective function can be represented as Eq. (3).

$$L(\emptyset) = \sum_i l(\hat{y_i}, y_i) + \sum_j \eta(f_j) \quad (3)$$

The first term of the equation is the loss function which is the summation of the prediction errors and the second term is the regularization to reduce the variance in the model [22]. The learning rate is one of the hyper-parameters of XGBoost that needs to be set optimally and should be set as low as possible to reach the best optimum but, this makes the computation slow. A trade-off between these two have to be maintained ensuring better performance at the same time. The learning rate η for the model used is set to 0.1, the regularization parameter γ to 1 and maximum depth to 2. The sub sample ratio of the training instances is set to 0.2.

5.6 *Gradient Boosting Classifier (GBC)*

Gradient boosting method is a more general implementation of XGBoost. This model also follows the step-wise additive model which generates learners during the process. Using the gradient descent optimisation, individual weak learner contributes to the ensemble [23]. It uses the loss function of the basic decision tree model as a means to minimize the prediction error of the overall model whereas XGBoost uses 2nd order derivative for approximation and advanced L1 and L2 regularization. Compared to XGBoost, Gradient Boost usually takes more time for getting trained.

5.7 *Gaussian Naïve Bayes Classifier*

Gaussian Naïve Bayes is an extension of the Naïve Bayes classifier to real-valued attributes by assuming that they have Gaussian distribution. The likelihood of the features of the data is assumed as in the Eq. (4).

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i-\bar{y})^2}{2\sigma_y^2}\right) \quad (4)$$

The model can be fit by finding mean and the standard deviation of the data points within each target label. At every sample point, the z-score distance between that sample point and each class mean is measured, which is the distance from class mean divided by the class's standard deviation [24]. This classifier doesn't require more training data and can handle both discrete and continuous data. The model is fast and its scalability with number of data points and predictors is very impressive. The hyper-parameter namely variance smoothing is tuned to *1e-01*. The variance smoothing value indicates the portion of largest variance of all the features which is added to variances for stability in calculation.

6. Results and Discussion

The dataset is imported into Jupyter notebook. The machine learning models namely Logistic Regression, k Nearest Neighbour, Support Vector Machine, Random Forest, eXtreme Gradient Boosting, Gradient Boosting, and Gaussian Naïve Bayes are used and evaluated using various metrics. Stratified cross-validation with 5 folds is performed in order to obtain the accuracy measure on the training and validation sets. Compared to k-fold cross-validation, stratified k-fold cross-validation will ensure that each of the folds get the same proportion of training examples with a given categorical value. This procedure makes each of the data points to pass through the training as well as validation phase. The held-out test set is completely hidden from the model and serves to check the generalization capability of the model. The scores of accuracy on the training set, validation set and the held-out test set obtained with each of the machine learning models used, is presented in Table 2.

Table 2. Accuracy scores for training, validation and held-out test sets for different models

ML Model	Training set	Validation set	Held-out test set
Logistic Regression	0.82	0.81	0.82
K-Nearest Neighbor	0.83	0.80	0.82
Support Vector Machine	0.84	0.80	0.88
Random Forest	0.83	0.78	0.82
XG Boost	0.88	0.78	0.80
Gradient Boost	0.81	0.78	0.84
Gaussian Naïve Bayes	0.81	0.81	0.90

Note: The hyper-parameter values were tuned to obtain maximum performance (discussed in Section 5).

Among the selected models, it is observed that Gaussian Naïve Bayes (GNB) followed by Support Vector Machine (SVM) models are showing significant performance on the held-out set. Logistic regression is also having good training set and validation set accuracy but not generalizing as good as SVM and GNB, on the held-out set. Moreover, checking merely the accuracy metric is not sufficient to evaluate a model and justify its generalizability. So, we have taken the following metrics into consideration to further gain confidence in evaluating the performance of the models used.

6.1 *Precision –Recall Curve – AUC*

Precision – Recall curve combines precision (Positive Predictive Value - confidence in the prediction of a positive class) and recall (True Positive Rate – fraction of all positive instances that the classifier correctly predicts as positive) in a single illustration. The larger is the area under the curve (AUC), the better is the model performance. PR curve is preferred to ROC curve as there is class-imbalance in the dataset being used, which can be better covered by a PR curve. Fig. 5 shows the P-R AUC for all the machine learning models considered except Gaussian Naïve Bayes. It can be noticed that AUC is more for SVM and logistic regression models whereas least in random forest classifier. In Fig. 6, the P-R curve for Gaussian Naïve Bayes model is shown and found to have the same measure as SVM and logistic regression. This makes it difficult to select between the models which further drives us to consider F1 score and log loss metrics to select the suitable model.

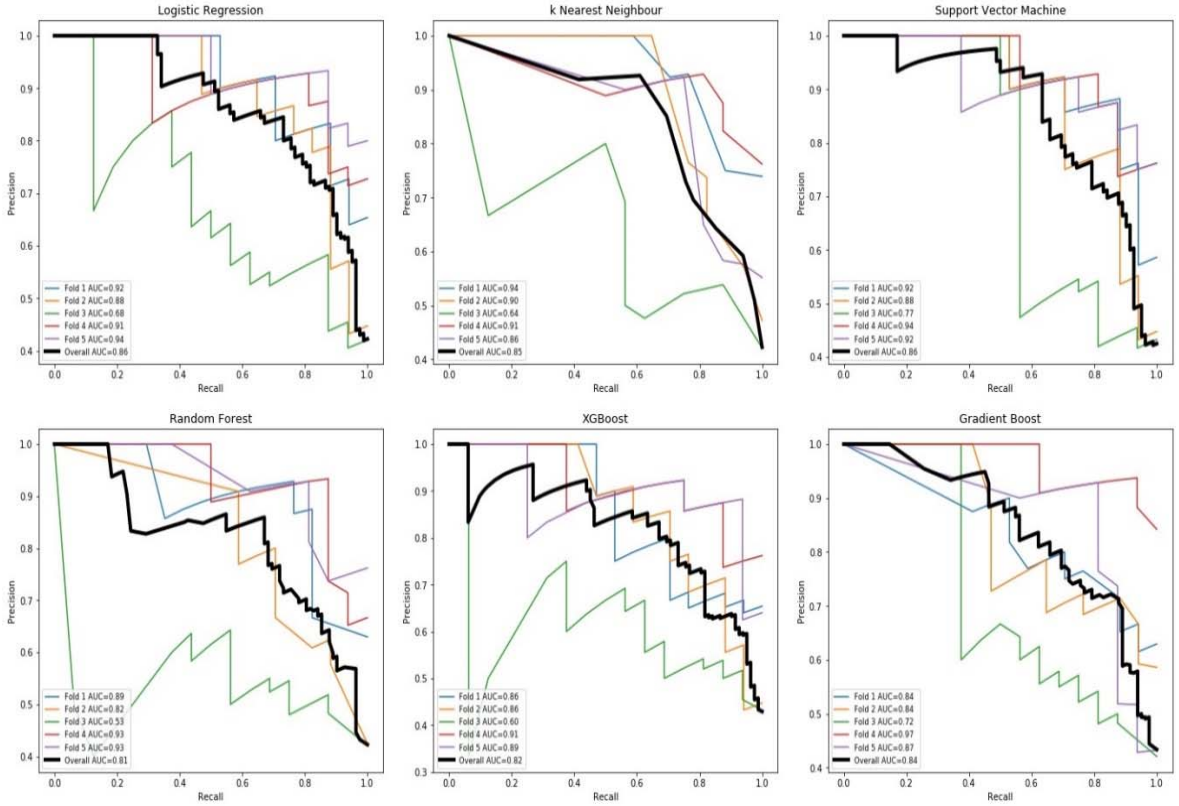

Fig. 5. P-R Curve AUC for the 6 models other than Gaussian Naïve Bayes

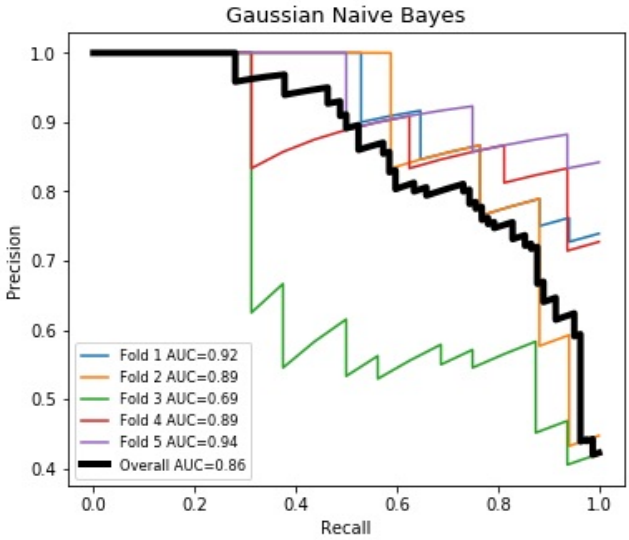

Fig. 6. P-R Curve AUC for Gaussian Naïve Bayes

6.2 *F1-Score*

Also called as F Measure or the F Score, F1 score explains the balance between the precision and recall. F1 score is more useful than accuracy score in the cases where class distribution is uneven. The F1 score is calculated using the formula in Eq. (5). The F1 score value lies between 0 and 1. The value towards 1 indicates a better score and thus better model.

$$F1 = 2 * \left(\frac{precision * recall}{precision + recall}\right) \quad (5)$$

The F1 scores of the machine learning models are represented in the form of a bar graph for better visualization in Fig. 7.

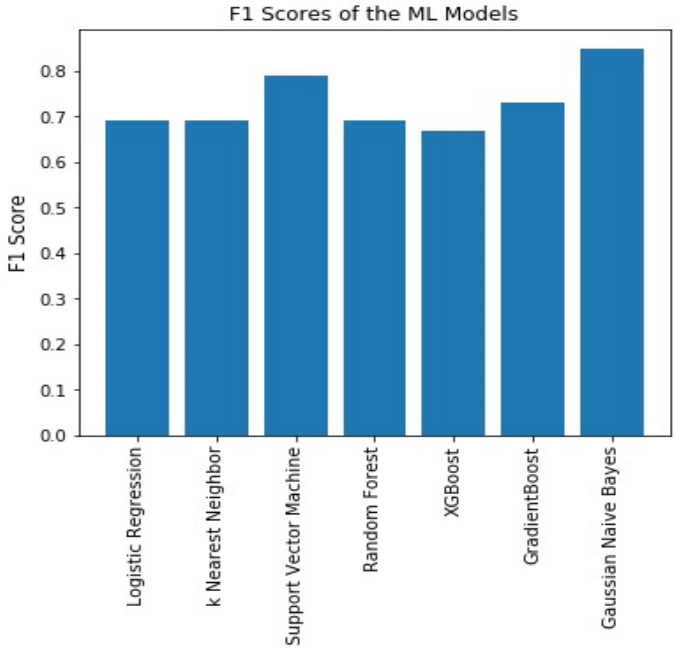

Fig. 7. F1 Scores of the machine learning models used

The F1 scores of Gaussian Naïve Bayes and SVM models are 0.85 and 0.79, respectively, having better scores compared to other models. Logistic regression's F1 score is 0.69 showing the uncertainty in its predictive capability.

6.3 *Log Loss (LL)*

Logarithmic Loss or Cross-Entropy Loss or simply Log Loss is a loss function and is the most important metric used in classification. It is a probability value that lies between 0 and 1. It quantifies the performance of a classifier model by penalising the false classifications. The representation of a log loss is given in Eq. (6).

$$LL_p(q) = -\frac{1}{N}\sum_{i=1}^{N} t_i \,.\, \log\big(p(t_i)\big) + (1 - t_i) \,.\, \log(1 - p(t_i)) \qquad (6)$$

In the above equation, *t* is the target label and *p(t)* is the predicted probability of the data point's label being Yes/positive class. It will favour the models that can more strongly distinguish the classes. This measure increases as the probability that is predicted, deviates from the actual target label. If the model predicts 0.009 when the observation label is actually 1, would result in high loss measurement. Fig. 8 shows the log loss of the machine learning models considered for the experiment and helps to take a decision on the selection of best performing model.

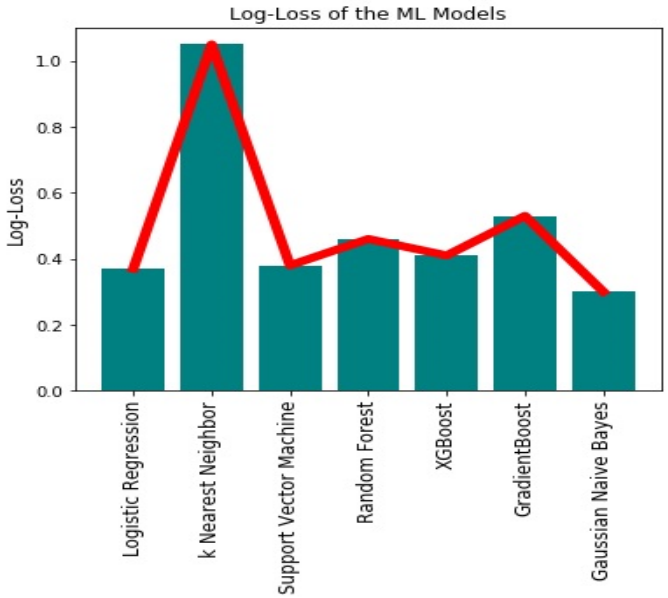

Fig. 8. Log Loss values of the machine learning models used

The log loss value for the Gaussian Naïve Bayes model at 0.30, is the least when compared to other models asserting its suitability to the problem at hand. The log loss values for SVM and logistic regression are almost the same at around 0.38.

The above results presented can help to infer the best performing machine learning model. Good amount of effort has been put into tuning the hyper-parameters of these models to achieve each model's maximum performance. Though most of the models perform well with respect to training data, they fail to retain the same performance with validation and held-out datasets. This also indicates that the model is overfitting with the training data and it would reduce the generalization capacity of the model. Held-out set accuracy scores enable us to understand how good a model is performing the prediction task on unseen data. Still, relying on these scores alone will make the evaluation task incomplete. Choosing a P-R curve AUC will help us to decide on a better model. As the classes are imbalanced, P-R curve would help us take better decision compared to ROC curve. F1 score is selected as another metric for evaluation as we care more about the positive class, that is, choosing palliative care in advanced stages of the Alzheimer's disease. Log loss will guide us to further evaluate goodness of the model based on the notion that, lesser the uncertainty in predictions, better is the model in its performance at the task. Taking all these metrics into consideration, we can conclude that Gaussian Naïve Bayes followed by SVM are the best performing models with better held-out set score, P-R curve AUC, F1 score and least log loss values. Most of the Alzheimer's disease related data will be in categorical format when it comes to analysing the behavioural traits, lifestyle, demographic and family relations. The inherent nature of Gaussian Naïve Bayes is to work well with categorical data that makes it suitable for the problem and standing out as the best performing model.

7. Conclusion and Future Work

The framework we have designed helps us to apply machine learning techniques in order to predict the decision to choose palliative care in the advanced stages of Alzheimer's disease. The data about the behavioural traits along with other relevant patient data was collected from the social media patient networks which provides a rich source of caregiver discussions. Rather than proctored surveys taken from the researcher or healthcare professional perspective, social media discussions contain dynamic data reported by caregivers. The caregivers directly deal with the patient and discuss about patient specific scenario, in the patient support groups, from their perspective. This helps the healthcare stakeholders to analyse the data and understand the trends in disease progression. Applying machine learning models to such data captured will assist them to take some informed decisions about treatment and hospice assistance required by the victims. The current experiment conducted finds Gaussian Naïve Bayes model, followed by SVM, as the best performing models in predicting the decision to choose palliative care in advanced stage of Alzheimer's disease. Future work is planned in the direction of studying the trends in patients' cognitive deterioration, rated with respect to different activities on various scales, using machine learning techniques.

Acknowledgements

We would like to acknowledge Ms. Meenakshi Banerjee, Research Scholar & Junior Consultant, Department of Clinical Psychology, National Institute of Mental Health and Neuro Sciences (NIMHANS), Bengaluru for providing the domain knowledge and assistance in the collection of relevant material, to progress with the research.

References

[1] "Alzheimer's Disease Fact Sheet", https://nia.nih.gov/health/alzheimer's-disease-fact-sheet, content reviewed: 22nd May, 2019, accessed on 28th August, 2020.
[2] Krista Hillis "Alzheimer's Statistics – United States & Worldwide Stats", https://www.braintest.com/alzheimers-statistics-throughout-the-united-states-and-worldwide/, accessed on 28th August, 2020.
[3] "Patient Care & Health Information – Diseases & Conditions – Alzheimer's Disease", https://www.mayoclinic.org/diseases-conditions/alzheimers-disease/symptoms-causes/syc-20350447, accessed on 28th August, 2020.
[4] Uddin, S., et al. (2019). Comparing Different Supervised Machine Learning Algorithms for Disease Prediction, BMC Med Informatics and Decision Making, 19, 281. https://doi.org/10.1186/s12911-019-1004-8.
[5] Caballe, N., et al. (2020). Machine Learning Applied to Diagnosis of Human Diseases: A Systematic Review, Applied Sciences. https://www.mdpi.com/2076-3417/10/15/5135.
[6] Suhad, A. S., et al. (2020). Alzheimer's Disease's Detection by using Deep Learning Algorithms: A Mini-Review, IEEE Access, 8. https://ieeexplore.ieee.org/document/9075205.
[7] Hill, N. L., Mogle, J. (2018). Alzheimer's Disease Risk Factors as Mediators of Subjective Memory Impairment and Objective Memory Decline: Protocol for a Construct-Level Replication Analysis, BMC Geriatrics, 18(1). https://bmcgeriatr.biomedcentral.com/articles/10.1186/s12877-018-0954-5.
[8] Moore, P. J., et al. (2019). Using Path Signatures to Predict a Diagnosis of Alzheimer's Disease, PLOS ONE, 14(9). https://doi.org/10.1371/journal.pone.0222212.
[9] Taeho, J., Kwangsik N., Andrew J.S. (2019). Deep Learning in Alzheimer's Disease: Diagnostic Classification and Prognostic Prediction using Neuroimaging Data, Frontiers in Ageing Neuroscience. https://doi.org/10.3389/fnagi.2019.00220.
[10] Javier, D. V. O., et al. (2019). Benchmarking Machine Learning Models for Late-onset Alzheimer's Disease Prediction from Genomic Data, BMC Bioinformatics 20(709). https://doi.org/10.1186/s12859-019-3158-x.
[11] Ben-Miled, Z., et al. (2019). Predicting Dementia with Routine Care EMR Data, Artificial Intelligence in Medicine. https://doi.org/10.1016/j.artmed.2019.101771

[12] Muller, S. F., (2003). Behavioral Disturbances in Dementia, Dialogues in Clinical Neuroscience, 5(1), pp. 49-59.
[13] Mega, M. S., et al. (1996). The Spectrum of Behavioural Changes in Alzheimer's Disease, Neurology, 46(1), pp.130-135. Doi:10.1212/wnl.46.1.130.
[14] Neugroschl, J., (2002). Agitation – How to Manage Change Disturbances in the Older Patient with Dementia, Geriatrics, 57(4), pp. 33-37.
[15] Max, K., Kjell, J., (2019). Feature Engineering and Selection – A Practical Approach for Predictive Models, CRC Press, ISBN: 1138079227.
[16] Recursive Feature Elimination, https://www.scikit-yb.org/en/latest/api/model-selection/rfecv.html.
[17] Juang, F., et al. (2017). Artificial Intelligence in Healthcare: Past, Present and Future, Stroke Vasc Neurol, 2, pp.230-243.
[18] David, W. H., Stanley, L., Rodney, X. S., Applied Logistic Regression, Third Edition, Wiley Series in Probability and Statistics, ISBN: 978-0470582473.
[19] Mucherino, A., Papajorgji, P. J., Pardalos P. M., k-Nearest Neighbour Classification, Data Mining in Agriculture, Springer Optimization and its Applications, 34, Springer, New York, NY. https://doi.org/10.1007/978-0-387-88615-2_4.
[20] Rohith, G., Support Vector Machine – Introduction to Machine Learning Algorithms, https://www.towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47.
[21] Liu, Y., Wang, Y., Zhang, J., (2012). New Machine Learning Algorithm: Random Forest, Information Computing and Applications, ICICA, Lecture Notes in Computer Science, 7473. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-34062-8_32.
[22] Malik, S., Harode, R., Kunwar, A., (2020). XGBoost: A Deep Dive into Boosting (Introduction Documentation). DOI: 10.13140/RG.2.2.15243.64803.
[23] Darren, C., Practical Machine Learning with H2O – Gradient Boosting Machines, O'reilly Publications.
[24] Rajeev, D. S. R., Yune-Sang, L., (2013). Smoothness without Smoothing: Why Gaussian Naïve Bayes is not Naïve for Multi-Subject Searchlight Studies, PLOS ONE. http://doi.org/10.1371/journal.pone.0069566.