IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 3, 2021, accepted March 10, 2021, date of publication March 17, 2021, date of current version March 23, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3066213

Computational Modeling of Dementia Prediction Using Deep Neural Network: Analysis on OASIS Dataset

SHAKILA BASHEER[1], SURBHI BHATIA[2], AND SAPIAH BINTI SAKRI[1]

[1]Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh 11671, Saudi Arabia
[2]Department of Information Systems, College of Computer Science and Information Technology, King Faisal University, Al Hofuf 36363, Saudi Arabia

Corresponding author: Surbhi Bhatia (surbhibhatia1988@yahoo.com)

The authors extend their appreciation to the Deputyship for Research & Innovation, Ministry of Education in Saudi Arabia for funding this research work through the project number PNU-DRI-RI-20-006.

ABSTRACT Alzheimer is a progressive disease and it is the most prevalent neurodegenerative disorder. It is believed that the people with mild cognitive impairment are at high risk of developing this disease. According to the annual report released by the Alzheimer's Association®2020, Alzheimer is the sixth leading cause of death in the United States. Thus, there is a need of educating people about this disease, reducing the risks by militating the necessary precautions to disseminate its affect by diagnosing it at early stages. It is also important to propose some recent advancement in this research which can help in early prediction of the disease using machine learning techniques. This paper intends to develop the novel algorithm by proposing changes in the designing of capsule network for best prediction results and making the model computationally efficient. The research is conducted on the Open Access Series of Imaging Studies (OASIS) dataset with dimensions (373 X 15) to diagnose the labels into two groups, as demented and non-demented. The novelty lies in conducting the in-depth research in identifying the importance of features, correlation study between factors and density of data showing status of factors by studying hierarchical examination of all the data points available using exploratory data analysis. Several optimization functions are conducted on the variables and feature selection is done to make the model faster and more accurate. The claims have been validated by showing the correlation accuracy at several iterations and layers with an admissible accuracy of 92.39%. The model is compared with state-of-art deep learning classifiers taken as benchmarks using different performance metrics. The ablation study is conducted on the proposed model using OASIS dataset to justify the predictions of the model.

INDEX TERMS Dementia, Alzheimer's disease, neural network models, machine learning, deep learning, convolutional neural networks, capsule networks.

I. INTRODUCTION

Dementia, a devastating illness that results in gradual loss of memory and other cognitive ability which is mostly identified in people more than 60 age groups. Alzheimer is a progressive disease that destroys memory and other important functions. Brain cell connections and the cells themselves degenerate and die, eventually destroying memory and other important mental functions [1]. An Alzheimer diagnosed person lives for four to eight years only. Few people live for twenty years as well because it completely depends on different factors. Various biological and neuropsychological studies discover that AD can be predicted at its early stage and useful to take treatment in an efficient direction. It starts from a specific subcortical region and increases to the cortical mantle with the passage of time. The most common effect of AD is memory loss and slows down the ability to do any task. It is found that MCI, a highly heterogeneous phenotypic spectrum, has very less considerable memory deficits than AD. These MCI may convert to AD in a study it was discovered that 10%-15% MCI patients converted to AD within a short span of time. So MCI needs to be taken care with special attention in order to stabilize the chance of AD [2]. The development of AD can be predicted several years before

The associate editor coordinating the review of this manuscript and approving it for publication was Juan A. Lara.

which are helpful in controlling the progress of AD. Biomarkers, magnetic resonance imaging (MRI), genetic data, cerebrospinal fluid, Positron emission tomography (PET) have attracted interest in identifying the early symptoms of AD dementia. MRIs do not involve ionizing radiation and are economical than PET and minimal invasive than cerebrospinal fluid (CSF). MRI provides multi-mode information for the brain's structure and function. MRI works successfully in distinguishing healthy people with AD survivors. MRI results can identify the sMCI (stable MCI) and pMCI (progressive MCI) [3]. These clinical and neuroimaging data have been used to extract feature information voxels, classify different groups and use it with several cognitive measures to produce support vector machine (SVM) based predictions, obtaining an area under Receiver operating characteristics (ROC) and authentication curve (AUC). Neuroimaging techniques are progressing very fast that makes it difficult to integrate large scale high dimensional multimodal neuroimaging data. Thus computer aided machine learning approaches are adopted for integrative analysis [4]. Linear discriminant analysis (LDA), linear program boosting method (LPBM), logistic regression (LR), support vector machine (SVM), SVM-recursive feature elimination have been used for early detection of disease. For any machine learning approach, architectural design and pre-processing steps should be predefined. For classification, the most popular four ways are feature extraction, feature selection, dimensionality reduction and feature based classification algorithm selection. Multiple optimization stages are required. In such methods features are extracted and selected via neuroimaging modality such as subcortical volumes, grey matter densities, cortical thickness, brain glucose metabolism and cerebral amyloid-beta accumulation in regions of interest. To work with the issues of machine learning, deep learning techniques are adopted which utilizes raw neuroimaging data to generate features via "on the fly" learning. It is good for large scale, high dimensional medical imaging analysis. CNN (convolutional neural networks), a deep learning method is used to study neuroimaging data for the early prediction of AD. A great deal of work has been made to develop a method to predict AD at its early stage in order to control its progress at the pre-symptomatic stage. A number of deep learning based research articles for AD have been studied and elaborated in related work. None of the previous researches claimed and emphasized on the use of a high performance modified capsule network for prediction of dementia on Open Access Series of Imaging Studies (OASIS) dataset [5], [6]. To develop an early prediction model of dementia based on the longitudinal data on OASIS dataset,[1] we did exploratory analysis of the dataset to find that the data has dimensions: 373 X 15. All these 15 factors are considered while analyzing and implementing various machine learning models. Exploratory data analysis (EDA) results and graphs are in OASIS EDA file. Analysis is done using various algorithms and their accuracy and other factors are considered. From the analysis, we see that none of the traditional methods give us accuracies up to 92.39%. We further observed that the data is hierarchical in nature and therefore, Capsule Networks (CapNets) were used for the task of prediction. However, due to the large number of features, we use a modified capsule network that does some optimization in the variables (W, K) and feature selection to make the model faster and more accurate. The results claim to achieve a considerable accuracy of 92.39% as compared with the other state-of-the-art machine learning and deep learning classifiers.

[1] https://www.oasis-brains.org/

A. MOTIVATION AND CONTRIBUTIONS

The motivation of the paper lies in early detection by extracting the complete visual features from the image. Second is to categorize all the images into proper labels. Finally to index these collection of MRI images and transforming into the series of text documents. In the recent scenario, many convolutional neural networks (CNN) works best with the images and has been used over a decade for labelling the images with achievable performance. But these methods only work well with large set of scaled dataset in order to develop the classifier for prediction tasks like in AD and works well in MRI images. The CNN can be trained on large image datasets where ConvNet can be used as a feature ex-tractor. There are drawbacks where CNN method is not applicable and so, CapNet known as highly structured machine learning designs can be used for image classification using deep learning technique. The faith of CapNet is because of its robust structure, it is able to access and fit in different data augmentation techniques and needs less dataset for training purposes. This paper aims to find the features by using CapNet which has capsules defined in its structure instead of neurons in neural network. The two output classes are defines as demented and non-demented in OASIS dataset. The input coming from CNN is fed in one capsule. The filtering of demented and non-demented features is done which basically depends on the features. The features are fed through the kernel principal component analysis (PCA), where large set of features are filtered into condensed features which are further passed into capsule for analysis. The auto encoder, which is a data analysis technique is used to encode the features using some labels so that the analysis of features are to be done in an easier and effective way. The classification is done to classify AD cases with demented and non-demented based on MRI images. Other than above, the paper proposes the implementation of modified capsule networks which will make the overall model as simple as possible to increase its reach in view of the humanitarian need to make it further applicable in hospital practice.

Following approaches are incorporated for this study:

- The hierarchical analysis is conducted on all the available data points for the prediction of dementia considering the 15 features in the OASIS dataset.
- EDA on the OASIS dataset is conducted representing the feature importance.

- Proposed the MCapNet model showing the implementation of the novel algorithm and how it serves advantageous with respect to original Capsule network.
- Analyze the accuracy of the MCapNet algorithm for predicting the disease on OASIS dataset
- Show the effectiveness of the proposed algorithm by comparing the results with the previously implemented state-of-art deep learning classifiers.

B. ARTICLE ORGANISATION

The remainder of the paper is structured as follows: Section 2 discusses the literature review. Section 3 gives the details on the proposed model shedding light on the methodology and the algorithm used to achieve the results. Section 4 describes the results and experiments done on the dataset showing exploratory data analysis with dataset. The experiments performed using different deep learning classifiers have been demonstrated. The accuracy results with the confusion matrix is displayed to show the effectiveness of the proposed technique. Finally, section 5 addresses a conclusion with future directions of research problem

II. RELATED WORK

This section explains about various existing techniques and tools to predict the early sign of AD and distinguish between a diseased and non-diseased person. All the existing tools and techniques are providing their benefits to people using neuroimaging data where images are retrieved to extract important features in prediction of AD [7], [8]. These features are clinically relevant information. For all these, various neural network techniques are applied to reveal texture and contour of hippocampus region extracted from MRI images. In this section we studied various deep learning based early prediction models. The authors addressed about the application of deep learning in the medical field. A thorough review of various algorithms of deep learning for diagnosis of Alzheimer's disease is done, in which it is found that the disease is a progressive brain disorder that destroys the brain memory gradually. It is a common disease in older age people which is caused by dementia. It has been obtained in most research papers that CNN is used as the most represented algorithm while dealing with brain image analysis. After a brief overview of various related papers for diagnosing AD, we come to know that the AD prediction at earlier stages can be increased by using advanced deep learning techniques in different dataset (ADNI, OASIS) combining to one [9]. Longitudinal studies investigate the incident dementia through deep learning techniques and predict the brain age based on MRI derived grey matter. This CNN based model successfully predicts brain age. The gray matter age prediction model is efficient in detecting brain age and chronological age. It uses logistic regression and cox proportional hazards which is significantly related to incident dementia. Grey matter density around amygdala and hippocampus is helpful in predicting age. And the age gap works as a biomarker which analyses the risk of dementia at its early stage [10]. The author explained that a combination of machine learning approach and stacked auto-encoder produced accuracies of up to 98.8% for classification and 83.7% for prediction for the conversion of mild cognitive impairment to AD. Machine learning approach is best suited for classification and SAE works well for feature selection. CNN and RNN deep learning approaches use neuroimaging data without pre-processing for feature selection produces 96% accuracy for classification and 84.2% for prediction for conversion of mild cognitive impairment to AD. A combination of multimodal neuroimaging with fluid biomarkers produces best classification results. This performance can be improved by obtaining some additional hybrid data types with some specific disease related features [11]. Li *et al.* discovered a deep learning model for automatic assessment of dementia severity which is applied to resting-state functional magnetic resonance imaging data (rs-fMRI). The author divides 133 AD patients with CDR (clinical dementia rating) and provides scores between 0.5-3. For extracting the features rs-fMRI is used with three-dimensional convolutional neural networks. This model is a very much acceptable in terms of accuracy [12]. The authors proposed another deep neural network model for identifying incident dementia using claims and EHR datasets. This is a predictive model based on deep learning which compares the effectiveness of performance to traditional machine learning models. This model is identified as an early clinical screening tool which monitors neuropsychology. EHR (Electronic health record) datasets from OLDW (Optum labs data warehouse) are very helpful in claiming the performance of this model. This model predicts the risk of AD 3-8 years before the onset of the disease. It is a feed-forward recurrent neural network model for predicting mild cognitive impairment and risk of incident AD. The EHR dataset includes data related to pharmacy, medical, laboratory, and enrollment of patients. This study is based on datasets of around ten years. The features such as age, gender and number of encounters are taken at the input layer. This model is expensive but provides accurate results [13]. The authors explored an effective deep learning tool for early prediction of AD which constructs classification methods applied on the brain. It was defined by Automated Anatomical labelling. AAL highlights gray matter images of the brain into 3D patches which are further used to train the deep belief networks. The Alzheimer's disease Neuroimaging Initiative (ADNI) data set is evaluated and it is obtained that the model is best suited for more challenging cases for classification of MCI [14]. There is no perfect method to early predict AD but researchers are working to provide various deep learning models to predict AD at its early stage by capturing different mild cognitive impairments and classify them to learn informative representation and temporal dynamics of longitudinal cognitive measures of individual subjects and combine them with baseline hippocampal MRI. The authors explained that a classifier with longitudinal data achieves better performance. The model is developed upon recurrent neural networks with long short term memory auto encoder

which learns compact and informative representation from longitudinal cognitive measures. The model collects data for baseline hippocampal MRI and learned data. It estimates the progress of MCI for AD dementia and applied to Alzheimer's disease neuroimaging initiative (ADNI). The model is well suited for prognosis purposes [15]. MCI (Mild Cognitive impairment) ultimately progresses to AD. Li *et al.* proposed a model for early prediction of AD dementia contingent on hippocampus MRI data. This model proves its efficiency by predicting MCI subject's progression to AD. It observes different patterns and adopts a classifier which ultimately cut-off at certain threshold points. This threshold point is basically helpful in observing the performance of the model. Hippocampal survey is a very important study for early prediction of AD. Different features for hippocampal study are shape, size and texture. These features are helpful in classifying different AD patients. This deep learning model is applied to the whole brain to monitor the progress of AD which is helpful in preclinical AD studies. This cognitive measure proves its performance in this study [16]. Ganguli *et al.* concluded that most MCI don't progress to AD but rather stable or may have chances to revert to normal. The study compared three groups who progress to AD, stable or revert to normal [17]. According to Spasov *et al.* all MCI are not the reason to convert AD. So the authors developed a model based on dual learning and an ad-hoc layer for 3D separable convolutions to predict most likely MCI to develop AD. This deep learning model takes MRI (magnetic resonance imaging), demographic, neuropsychological and APOe4 genetic data as input measures. The model is multi-tasking which is able to predict the chances of MCI to AD and AD itself. It works on selected parameters to avoid the situation of data-over fitting. The warp field works as an additional predictive value for AD prediction. It distinguishes the MCI survivor developing AD within 3 years to the patients with stable MCI over the same period of time. This model is flexible for other imaging modalities as well such as PET etc. [18]. Qureshi *et al.* discover a novel computer-aided diagnosis system which utilizes feature ranking along with genetic algorithms for analyzing structural magnetic resonance imaging data. The model can successfully predict the conversion of MCI to AD between one to three years before clinical diagnosis. The features are ranked as per their t-test scores and genetic algorithm design to obtain optimal feature subset. Fisher criteria works as an objective function in genetic algorithms. This model proves its efficiency in distinguishing stable MCI to progressive MCI and reduces the dimensionality of the input vectors to low dimensional space [19]. Ding *et al.* also proposed deep learning based model to predict final diagnosis of AD and MCI by using fluorine 18 fluorodeoxyglucose (FDG) PET of the brain. It compared the performance with conventional radiologic readers and collected data of 40 patients from 2006 to 2016. The final diagnosis and results indicate that the model is sensitive, specific with major characteristics as receiver operating system, saliency map and t-distributed stochastic neighbor embedding. Saliency map is helpful in detecting areas of interest. The model is 82% specific, 100% sensitive and good enough for early prediction [20]. With an improvement on study of early diagnosis of AD, Liqin *et al.* asked to focus on palliative care for dementia patients as palliative care may help in reducing the risk of mortality. In this study the author collects data for prediction of 6-month, 1 year and 2 year mortality. The model is a recurrent neural network which collects demographic data of patients for 2011 to 2017. The data model is trained with 24229 datasets and validated with another 2692 dataset which was analyzed from September 2018 to Dec 2019 [21]. To check the progress in AD, authors proposed to develop the minimal RNN model to forecast AD progression. The minimal RNN has less parameters than other RNN models, such as the long short-term memory (LSTM) model, so it can be slight prone to over fitting. Although RNNs are usually trained using feature-complete data, ''preprocessing'' and ''integrative'' approaches are considered to deal with missing data. The data is applied from the TADPOLE competition, comprising longitudinal data from 1677 participants with an achievable accuracy [22]. Liu *et al.* designed a deep learning architecture, which contains stacked auto-encoders and a softmax output layer, to control the bottleneck and support the diagnosis of AD and its leading stage, Mild Cognitive Impairment (MCI). Compared to the previous work, this method is highly proficient of analyzing multiple classes in one setting, and requires slighter labeled training samples and reduce domain prior knowledge [23]. Park *et al.* works on integration of gene expression and DNA methylation dataset and addressed a deep learning-based model that can forecast AD using large-scale gene expression and DNA methylation data. A novel but simple approach to reduce the number of features based on a differentially expressed gene and a differentially methylated position in the multi-omics dataset. This article demonstrated that integrating gene expression and DNA methylation data could improve the prediction accuracy [24]. According to a study by Karlekar *et al.* changes in language is also a sign for predicting AD at its initial stage. Natural language processing based models classify different linguistic characteristics from Dementia Bank dataset. The authors combines three neural network classification models (CNN, LSTM-RNN and CNN-LSTM) to challenge the performance of said model. The analysis depends upon activation clustering for automatic pattern discovery and first-derivative saliency technique which rediscovers previous language patterns of AD survivors [25]. The classification accuracy of 95.59%, with a sensitivity of 97.06% and a specificity of 94.12% with leave-one-out cross validation (LOOCV). The paper forecasts progression in AD and demonstrates the generative models capable of sampling conditional probability distributions. The authors emphasized on the utility of cognitive scores as a measure of disease activity [26]. This method minimize data redundancy on brain data using two methods such as Common spatial DCADx (DCADx-CSP) and brain functional network DCADx (DCADx-BFN) DCADx-CSP obtains 82.5%

accuracy and DCADx-BFN obtains 88.7% accuracy. The authors applied CBIR (Content based image retrieval) for preclinical observation with image classification tools such as CNN and capsule network using MRI scan data on ADNI dataset. A pre-trained 3D autoencoder is used to improve the accuracy as compared to state of art CNN model. Model gave an achievable accuracy but lacks in term of humanitarian factors [27]. Capsule networks encapsulate features into groups of neurons using NLP technique and HMC classifies the entire hierarchy of data at once. This is a global approach, but not yet implemented and capsule network in HMC is not adopted yet [28]. Ivan Izonin *et al.* proposed a new method for solving multiple linear regression task via a linear polynomial by proposing constructive formula. The novel method is based on the repetition of training outcomes and the lack of debugging and parameter selection procedures, allows synthesizing linear polynomial for complex models that use various non-linear extensions of SGTM inputs while preserving the accuracy of their operation. The proposed approach can be used in the fields of medicine, economics, materials science, service sciences etc., for fast and accurate solution of regression or classification tasks with the possibility of easy interpretation of the result [29]. Nataliia Melnykova *et al.* proposed non iterative method to personalize the treatment by investigating the individual patient characteristics by applying the Machine learning and Data mining methods. It is proposed to classify persons by conditions, to determine deviations of parameters from the normative parameters of the group, as well as the average parameters. This model is to create a personalized approach for treatment based on long -term monitoring, According to the result of the analysis, it becomes possible to find the medicaments treatment according to the personal characteristics of a patient [30]. The previous researches and the key factors employed for classifying the disease for comparison with the proposed model has been given in Table 1.

TABLE 1. Previous researches for comparison.

Paper	Technique/ Methods	Datasets	Key factors
Hierarchical multi-label classification of text with capsule network [31]	Capsule network for hierarchical multi-label classifier`	BlurbGenreCollection and Web of science dataset	Capsule networks encapsulate features using NLP technique and HMC classifies the entire hierarchy of data at once. This is a global approach, but not yet implemented. Capsule network in HMC is not adopted yet.
Computer aided dementia diagnosis based on hierarchical extreme learning machine [32]	Hierarchical extreme learning machine (H-ELM)	ADNI (AD neuroimaging initiative) dataset	This method minimize data redundancy on brain data using two methods such as Common spatial DCADx (DCADx-CSP) and brain functional network DCADx (DCADx-BFN) DCADx-CSP obtains 82.5% accuracy and DCADx-BFN obtains 88.7% accuracy.
Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks [33]	Deep CNN using brain MRI data analysis	MRI images	The model identifies different stages of AD. The accuracy of the proposed model is 93.18% with 94% precision, 93% recall and 92% f1-score.

A. JUSTIFICATION OF PROPOSED WORK

The objective of proposed model is to early predict AD, so that proper care and treatment can be given to the patient. The proposed model is modified capsule networks or Modified CapsNets (M-CapNets) which is more efficient than CNN. To authenticate the validity of proposed model and to train the model, OASIS data set is used. M-CapsNets overcome the issues of CNN and provide promising results for medical image analysis. This novel model will benefit the people living inside the kingdom and research in the medical community to take better decisions in early stages of dementia. The related work study discovers that most of the work on AD is done under CNN networks. But there are various limitations of CNN, such as lack of spatial information and pooling function. The basic Capsule network has certain limitations as far as computational factors are compared with respect to time and space complexity. M-CapNets as proposed in this paper will certainly be advantageous as compared with the previous researches with deep learning architectures in terms of computational power and accuracy. M-CapNets provides better hierarchical relationship values and uses feature detection method to increase the run time of the model. Kernel PCA method is used for object detection. Thus, this novel technique intends to make this model as simple as possible to increase its reach in view of the humanitarian need to make it further applicable in hospital practice. The M-CapNets achieves 94.39 accuracy of prediction.

III. PROPOSED METHODOLOGY

The focus of this paper highlights the utilization of classification methods by application of query through feeding images in a type of image retrieval system, using the CNN method and the proposed technique of using CapNet. The general flow chart is given in Figure 1. It works by extracting images or taking some sample images. The image preprocessing techniques are applied such as filtering, restoration, removing background noise, segmentation to

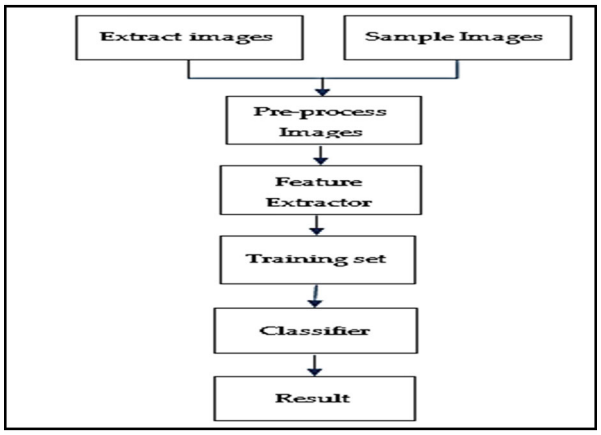

FIGURE 1. Flowchart of the proposed model.

extract relevant information [34]. The feature extraction techniques are applied for extracting patches from the image and thereafter, mathematical transformation is done for generating feature vectors. Finally the model is trained and built. For all the weights and the bias from labeled cases, training a model means determining good values. A machine learning algorithm generates a model in supervised learning by analyzing several examples and try to find a model that minimizes loss, which is known as minimization of risk/error.

A. DATA PREPARATION

The data used for preparation of the paper is taken from OASIS database (https://www.oasis-brains.org/) [35]. The dataset consists of MRI and PET images of clinical patients collected over the period of 15 years. The study was done on longitudinal MRI dataset consisting of 150 subjects for exploratory data analysis. The details on the dataset is given in the experimental results sections.

B. DESIGNING OF CAPNET

CNN works on image data and detects features of an image to recognize objects with information. In CNN, the layers detect only edges of an object and layers that are deeper can detect complex features of an object [36]. The CNN based deep learning model basically uses all the learned features to make final predictions. The major flaws of CNN are absence of spatial information and pooling function. In max pooling operation, only important information found from the most active neurons are to be gathered to the next layer. Due to this, some important spatial information gets lost between the layers. Therefore, to handle the above issues of CNN, we use a more refined form of CovNets called CapNet or Capsule Networks. Capsule Network is just a variant of a CNN. CapNets were devised to handle hierarchical modeling problems and are the most suitable for this research problem. CapNets does not resembles the Pooling layers used in CovNets. Because of pooling in CovNets, it is used to reduce the details and increase the speed of algorithmic runtime. However in CapNets, pooling layer considers the minor details into account which is based on CapNet (concept of inverse rendering). The beauty of CapNet lies in the structure represented by nesting of convolutional layers. The architecture is shown in below Figure 2.

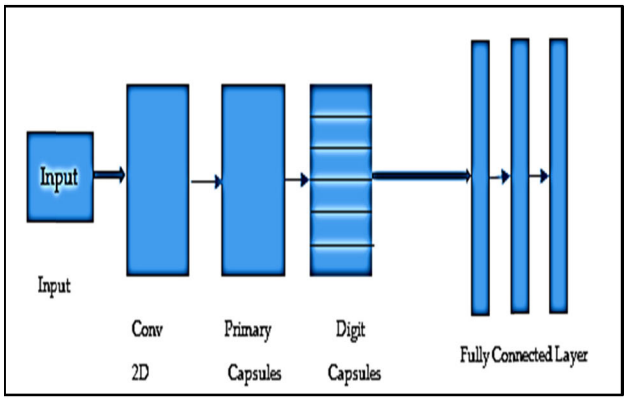

FIGURE 2. Basic Architecture of CapNet model.

The two important functions of CapNet are Routing algorithm and Squashing [37]. The routing is referred to as coupling between capsules and squashing is done on the DigitCaps layers. During training, the activation vector of the correct digit is masked out and this activity vector is used to reconstruct the input image using fully connected decoder.

C. PROPOSED MODEL

The proposed method uses modified capsule network, which takes the data as an input, makes the parent vector by considering the various features available in the OASIS dataset. The CNN with RELU activation function is that in which the data is fed which gives the transformation vector w. This w along the output from various capsules is used using the logic of dynamic routing for us to get the feature. The KPCA is used to do a feature extraction before passing the memory from one capsule to another. Squash is a special type of activation function used in Capsule Networks to normalize the magnitude of vectors, instead of the scalar elements themselves. The findings from these squash functions inform us how data can be routed through different capsules that are equipped to learn various concepts. This helps in making the system more computationally efficient. Squash is a mathematical step in which a 3D matrix is squashed into an array or a column vector. Squashing is mathematically defined as represented in Equation 1.

$$as, \quad a \leftarrow \frac{||a||^2}{1+||a||^2} \frac{a}{||a||} \tag{1}$$

The proposed model is illustrated in Figure 3.

The modifications in the proposed Capsule Network are highlighted. The steps are mentioned numerically in the algorithm and explanation of each step is written below:

- The standard capsule network feeds the entire output from the previous capsule to the next. This does not

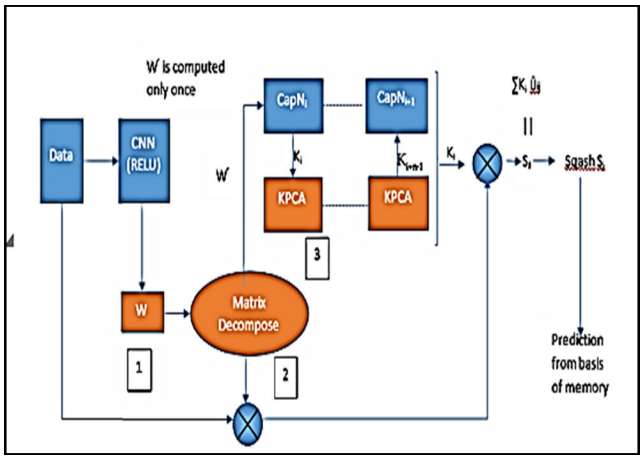

FIGURE 3. **Proposed Model showing Modified Capsule Network Architecture.**

affect the accuracy, however, this takes a big toll on the computational efficiency of the model. To prevent this, we use an additional feature extraction layer – KPCA – to do a feature extraction before passing the memory from one capsule to another.

- The standard capsule network uses a multidimensional matrix as a transformation matrix. In the modified method, we do a matrix decomposition of this using SVD.
- The standard capsule network calculates the transformation at each capsule level, in our modified model, we calculate it only once and propagate it across the network.

1 ALGORITHM

We have a parent vector u_i which is translated and rotated by a decomposed transformer matrix W'_{ij} into a unit vector defined as $\hat{u}_{j|i} = W'_{ij}u_i$.
Say, there are m capsules, the sum of predictions from m_n capsules are fed into the m_{n+1} capsules, with a coupling coefficient k_{ij}.

$$s_j = \sum k_{ij}\hat{u}_{j|i}$$

The coupling coefficient is the most important factor in this algorithm. The pseudocode of CapNets is given below,
1: Ingest Data (D)
2: Calculate the parent vector, u_i (Contains the encoded values of features).
3: Define two sub-capsules, i meaning Dementia is present and j meaning Dementia is absent.
4. Run basic CovNet with ReLU activation function to generate a vector, say a.
5: Squash the vector a into a vector to fit our feature set, squashing is done using the following formula:

$$a \leftarrow \frac{||a||^2}{1+||a||^2}\frac{a}{||a||}$$

6: Each squashed vector is a unit vector and is used in a capsule along with a transformation matrix W_{ij} to explain the hierarchical relationship between the classes mentioned in the respective capsule.
7: Matrix decomposition of W_{ij} to get W'_{ij}
8: W'_{ij} is computed only once.
9: From each capsule, the results are forwarded to the next capsule along with a coupling coefficient k_{ij}.
10: This coupling coefficient is calculated using the activation function of the CapNet. Therefore, $\sum k_{ij} = 1$.
11: Now, CapNet gives great results due to the dynamic routing of the results across the layers. The algorithm for dynamic routing is given below:
A: Define a (Unit Vector)
B: Define r (Iterations, Scalar)
C: Define l (No. of layers, Scalar)
D: foreach capsule n in layer l, capsule m in layer $(l+1)$ do *memory* <- 0
E: for r do:
E1: foreach capsule n in l do *ki* <- *softmax (b,n)*
E2: for each *ki*, do *ki'* <- *KPCA (ki)*
E3: foreach capsule m in *(l+1)* do *sj* <- $\sum k'_{ij}\hat{u}_{j|i}$ ($\hat{u}_{j|i}$ is calculated using the decomposed W'_{ij})
E4: for each *sj* do *sj'* <- *KPCA (sj)*
E3: foreach capsule m in *(l+1)* do a <- *squash (sj')*
E4: foreach capsule n in l, m in *(l+1)* do *memoryij* <- *memoryij* $+\hat{u}_{j|i}$. v
F: Return v.

IV. RESULTS AND DISCUSSIONS

The analysis of early prediction of dementia is done on OASIS MRI longitudinal dataset which is a freely available dataset for the scientific community for the aim to collect neuroimaging, clinical, cognitive and biomarker datasets. It is a multi-modal dataset for clinical neuroscience, hosted by central.xnat.org. It is an easily accessible platform for neuroimaging, clinical and cognitive research.

A. DETAILS ON DATASET

OASIS dataset consists of a longitudinal collection of 150 subjects aging from 60 to 96 years old [38]. Each subject was scanned on two or more visits, separated by at least one year for a total of 373 imaging sessions. For each subject, 3 or 4 individual T1-weighted MRI scans obtained in single scan sessions are included. The subjects are all right-handed and include both men and women. In this analysis, 72 of the subjects were characterized as non-demented throughout the study. 64 of the included subjects were characterized as demented at the time of their initial visits and remained so for subsequent scans, including 51 individuals with mild to moderate Alzheimer's disease. Another 14 subjects were characterized as non-demented at the time of their initial visit and were subsequently characterized as demented at a later visit. Table 2 has the collection of number of diagnosed

TABLE 2. Labels for diagnosis on OASIS dataset.

Diagnosis	No. of Patients/Individuals	Gender		
Row Labels		**F**	**M**	**Total**
Converted	37	24	13	37
Demented	146	60	86	146
Non-demented	**190**	129	61	190
Grand Total	**373**	**213**	**160**	**373**

people into three categories demented, non-demented and converted.

The dataset available with us has various features such as MR Delay, Gender, Dominant hand, Age, Education, SES, MMSE, CDR, ASF, eTIV and nWBV [38]. The sample characteristics are as clinical dementia rating (CDR) displayed in Table 3.

TABLE 3. Few samples and features taken for analysis.

	CDR 0	CDR 0.5	CDR 1
Number	86	51	13
Female/male	60/26	21/30	7/6
Age (years)	(60–93)	(62–90)	(61–96)
Education (years)	(8–23)	(6–20)	(8–20)
MMSE	(27–30)	(17–30)	(19–30)
Prescriptions (n)	(0–9)	(0–11)	(0–7)
Systolic BP (mmHg)	(98–192)	(118–188)	(90–188)
Diastolic BP (mmHg)	(50–100)	(58–98)	(60–88)
Reported HBP (%)	54.6	46	53.3
Diabetes (%)	9.3	14	13.3

For best prediction of dementia using the available data, we are required to do hierarchical analysis of all the data points available. The various features such as MR Delay, Gender, Dominant hand, Age, Education, SES, MMSE, CDR, ASF, eTIV and nWB are taken for the prediction of dementia using the available data, the hierarchical analysis of all the data points available are performed using the Modified Capsule network.

The features in the analysis are explained as under in Table 4.

Figure 4a shows the classification of number of counts for demented and non-demented groups for their first visit. Figure 4b, shows total number of demented and non-demented groups only, as converted people are now added with demented groups. Converted groups are those patients who were converted into dementia state in later visits. As shown in Figure 4c that male and female counts are almost equal in demented group. But female count is more than the male count in non-demented group.

B. EXPLORATORY DATA ANALYSIS ON DATASET

The four basic python libraries are used such as numpy, panda, seaborn and matplotlib are used for conducting EDA. After applying inputs, the shape of the data is checked.

TABLE 4. Details of the features.

Feature Name	Description
MR Delay	Delay between the first and second MRI Scans.
Gender	Male/Female
Orientation of hand movement basis	Right Handed/Left handed
EDUC	Years of Education
SES	Socioeconomic Status (It actually correlates the dementia with the socio economic conditions, socioeconomic status as assessed by the Hollingshead Index of Social Position and classified into categories from 1 (highest status) to 5 (lowest status).
MMSE	Mini-Mental State Examination score (range is from 0 [worst] to 30 [best]) (basic function balance, sense of height, balance, taking all these considerations taken a scale) (Folstein, Folstein, & McHugh, 1975)
CDR	Clinical Dementia Rating. (0= normal, 0.5= questionable, 1= mild, 2= moderate, 3= Severe, 4= profound and 5= Terminal.)
ASF	Atlas scaling factor (unit less). Computed scaling factor that transforms native-space brain and skull to the atlas target (i.e. the determinant of the transform matrix) (atlas is the highest bone of the body, this atlas scaling factor shows your brain is at the right position)
eTIV	Estimated total intracranial volume (cranium within our skull and the brain volume)
nWBV	Normalized whole brain volume, expressed as a percent of all voxels in the at-las-masked image that are labeled as gray or white matter by the automated tissue segmentation process.

The data contains 373 rows and 15 columns. This data have information regarding number of visits in hospital by the patient. As shown in Figure 5a, dementia is more common in men's than women.

In violin plot blue portion is for men and yellow is for women. CDR by gender can be viewed in figure 10. Clinical dementia ratings can be defined between 0-5. In this 0 = normal, 0.5 = questionable/ very mild, 1 = mild, 2 = moderate, 3 = Severe, 4 = profound and 5 = Terminal. As shown in Figure 10 violin plot of gender by CDR, normal females are more than male, questionable males are greater than female, mild rate are more prevalent in male, severe cases in male and female are almost equal, so it is found that dementia is gender independent. In Figure 5b, dementia is observed as per age and discovered that the chance of dementia are more at an age of 70-80 and of course age plays a significant role.

In box plot shown in Appendix presents the oscillation values for density of data and shows status of various factors. The dark part in box plot shows the maximum density at some points. The correlation sown in figure 6 between gender and CDR is also analysed discovering that gender plays no role in correlation matrix. The diagonal values in correlation matrix are always one, as it presents a strong correlation. The correlation between gender and CDR gives the value is 0.57 which is higher than moderate correlation. It shows that it is somehow gender dependent. The results are shown

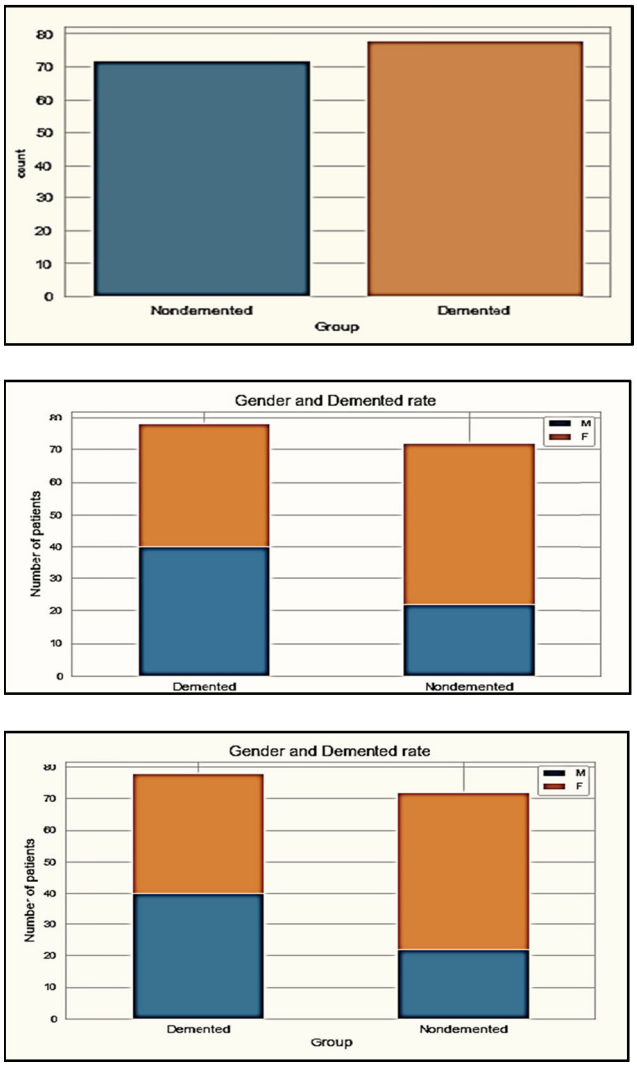

FIGURE 4. (a) Demented and non-demented groups and Converted Groups, (b) Number of counts for demented, Non-Demented, (c) Gender and demented rate.

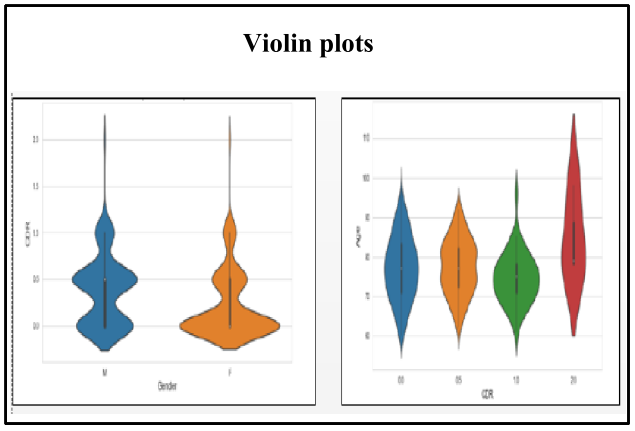

FIGURE 5. (a) Demented with respect to gender, (b) Dementia with respect to age.

in appendix A for correlation matrix. The dark part in box plot shows the maximum density at some points.

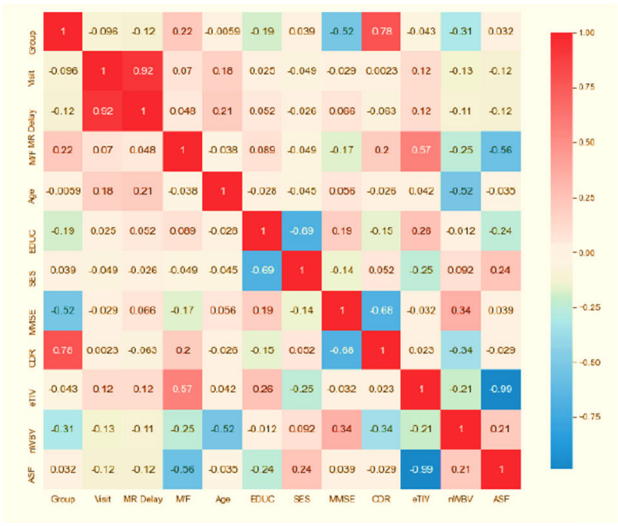

FIGURE 6. Correlation plot matrix.

The correlation matrix includes numeric values of correlation between various features. The correlation between gender and CDR, gives value 0.2 that is a very weak correlation and discovers that gender plays no role in correlation matrix. The diagonal values in correlation matrix are always one, as it presents a strong correlation. The correlation between gender and CDR gives value as 0.57 which is higher than moderate correlation. It shows that it is gender dependent.

C. EXPERIMENTAL SETTING

This experimental study is divided into two parts. Firstly, a taxonomic hierarchical data taken from OASIS dataset is used and analysis of the entire dataset is completed based on the EDA. The modules used for analyzing the dataset was Numpy (For matrix analysis), Seaborn (for advanced plotting), pandas (for data ingestion) and Matplotlib (for basic plotting), sklearn and tensorflow (for Machine learning tasks). The Inter i5 8th Gen processor was used with 16GB RAM and CentOS operating system for achieving the results. Secondly, the Amended Capsule network was applied. Considering total 14 features Excluding ID – Column A of OASIS dataset. The total Iterations in amended capsule networks were 1000 with number of Capsule Layers were taken to be 50. The number of features inside capsule is categorized as demented or non-demented. The variation of accuracy with the number of layers and number of iterations is shown in Figure 7 represents the accuracy at each iteration and Figure 8 represents accuracy at different layers.

1) MODEL BASED FEATURE REPRESENTATION

To build an early prediction model of AD dementia based on longitudinal data, we first train the dataset to learn compact representation and encode the dynamics of longitudinal measures for each subject. This work contains hyper-parametric classifiers from Naïve Bayes Models, Support vector machine (SVM), Decision Trees, XGBoost, Gradient Boost, Ensemble, Adaboost, Random forest models, Light gradient boosting framework machine LGBM [39]–[41].

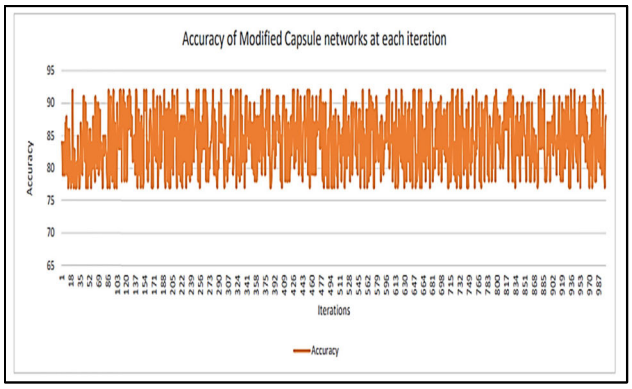

FIGURE 7. Correlation Accuracy of modified capsule network at different iterations.

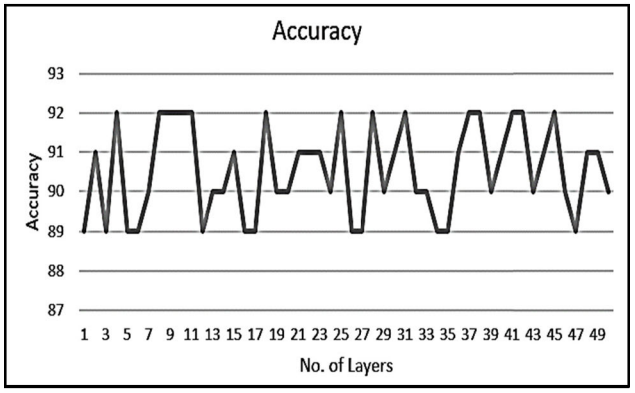

FIGURE 8. Correlation Accuracy of modified capsule network at different layers.

Support vector machine classifier algorithm works on developing linear decision boundries to classify multiple classes. SVM are supervised learning models with associated learning algorithms which analyze data used for classification and regression analysis. Decision tree creates classification or regression models in the form of a tree structure which breaks down a dataset into smaller and smaller subsets and subsequently decision tree is incrementally developed. XGBoost and Gradient Boost a special type of Ensemble Learning technique which perform combination of various weak learning and strong learning and focus on its predecessor's faults to improve its accuracy. Bagging: The accuracy of classification and regression tree can be improved through bagging and also known as ensemble method. Adaboost classifier iteratively trains the machine learning model with the help of training set which are based upon accurate prediction of precious training. The iteration ended till whole training data fits with minimal error or no error. The random forest classification algorithm includes many decisions trees and support bagging as well as feature randomness while building each individual tree to make an uncorrelated forest of trees. Its prediction results are more accurate than that individual tree. LGBM classifier works on tree based learning algorithms LGBM has faster training speed and higher efficiency which support parallel and GPU learning. It is capable of handling large-scale data. Naive Bayes: The Naive Bayes classifier works on Bayes' theorem of probability and produces very accurate classification output with a reduced training time when compared to conventional supervised or unsupervised learning algorithms. It outperforms conventional classifiers in terms of training speed and classification accuracy.

Figure 9 showing the bar chart demonstrates that the proposed technique using MCapNet performs better than the other machine learning classifiers in terms of accuracy. The parameters such as sensitivity rate and error rate is also shown as comparison with the other state-of-art classifiers. The model shows the highest accuracy achieving 92.39% and lowest error rate of 7.61%.

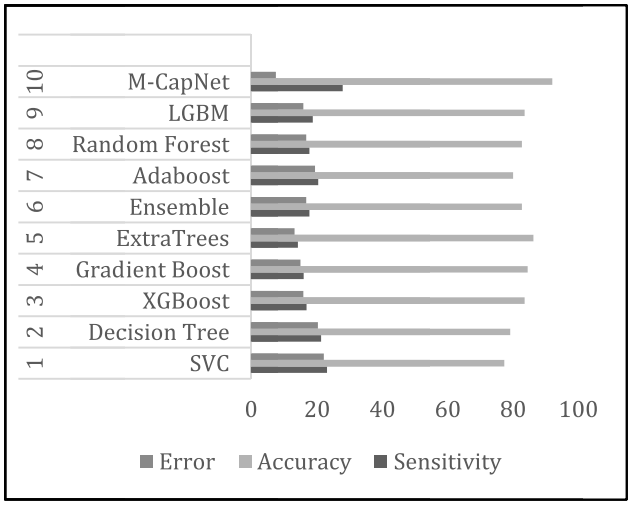

FIGURE 9. Accuracy, Error rate and sensitivity rate calculated on OASIS dataset.

The recall, precision and f-score is calculated for all identified methods and are shown in Figure 10.

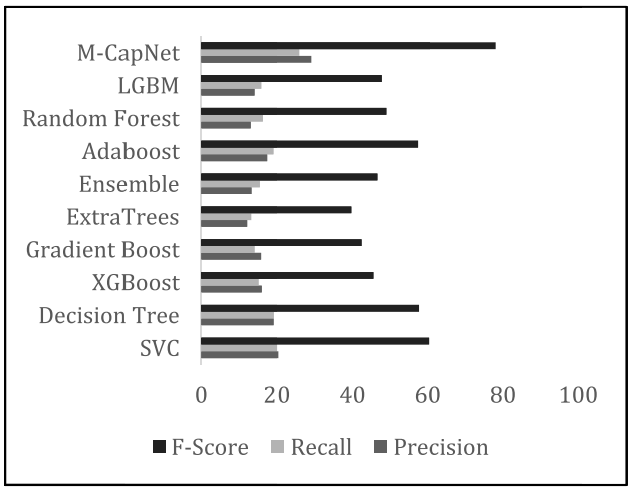

FIGURE 10. Recall, Precision and F-Score of the methods.

An ROC curve (receiver operating characteristic curve) is a graph showing the performance of a classification model at

all classification thresholds as shows in Figure 11. It demonstrates the behavior of the model, showing the time taken to reach the threshold. The threshold value is 1 showing that all the classifiers taken for comparison with the proposed technique are valid. However, the area under the proposed technique (MCapNet) is highest, showing that the proposed model converges quickly giving the maximum accuracy. This curve plots two parameters; True Positive rate on the X axis and the False Positive Rate on the Y axis. The True Positive Rate (TPR) and the False Positive Rate are identified at varied points, and the same is plotted on a graph. The True Positive Rate (TPR) is otherwise called as sensitivity and the False Positive Rate (FPR) is otherwise called as specificity.

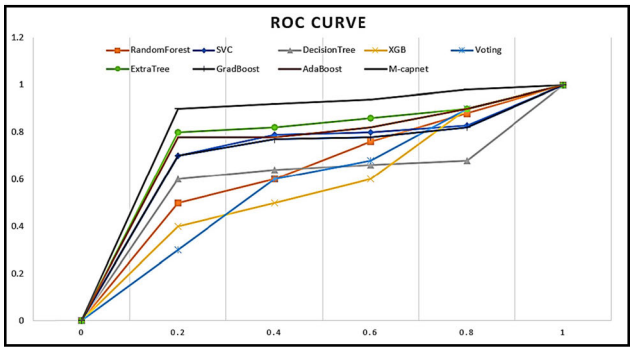

FIGURE 11. ROC Curve.

It can be calculated as shown in Equation 2 and Equation 3. Here TP is true positive that contains the relevant information and FP means irrelevant information.

$$\text{TPR} = \text{TP}/(\text{TP} + \text{FP}) \tag{2}$$

$$\text{FPR} = \text{FP}/(\text{TP} + \text{FP}) \tag{3}$$

Figure 12 is showing the runtime (in seconds) for the various classifiers. The models was run Inter i5 8th Gen processor without any graphic card. All the models were run independently on the OASIS dataset in order to observe the run time of each model. It gives the indication of time complexity. The rest of the models are all lagging behind by the time difference varying between 0.3 (lowest) to 2.3 (highest) seconds.

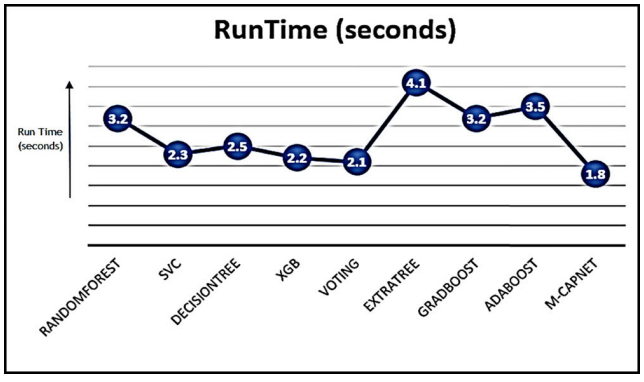

FIGURE 12. Run time graph.

The outcome of the result shows that the proposed technique developed using MCapNet takes the least time of 1.8 second to run, making our model computationally efficient and fastest.

D. ABLATION STUDY

In order to understand the component's contribution to the overall system, an ablation analysis studies the efficiency of an artificial intelligence (AI) system by removing certain elements. As the proposed model is showing the analysis on OASIS dataset that employs different features, it is essential to evaluate each of the features quantitatively. This will allow us to investigate not only the test accuracy but also the trade-off between the accuracy. We evaluate the test dataset divided into four different sets as A, B, C and D. Set A includes all the features given in the OASIS dataset, Set B includes all the features excluding feature age. Set C excludes two features, i.e. age and gender and Set D excludes two features, i.e. Gender and SES. It has been observed that when all the features are taken the accuracy is showing the result of 92.3%. But when the feature age has been excluded, the accuracy has been dropped to 71.1%, further removing features age and gender, the accuracy dropped to 69.9%. However removing features such as Education and SSE doesn't show any reduction in the accuracy. This validates that age is highly dependent feature and must be considered for predicting the disease. The graph showing the trade of between the various features with respect to accuracy and F1score is represented in Figure 13.

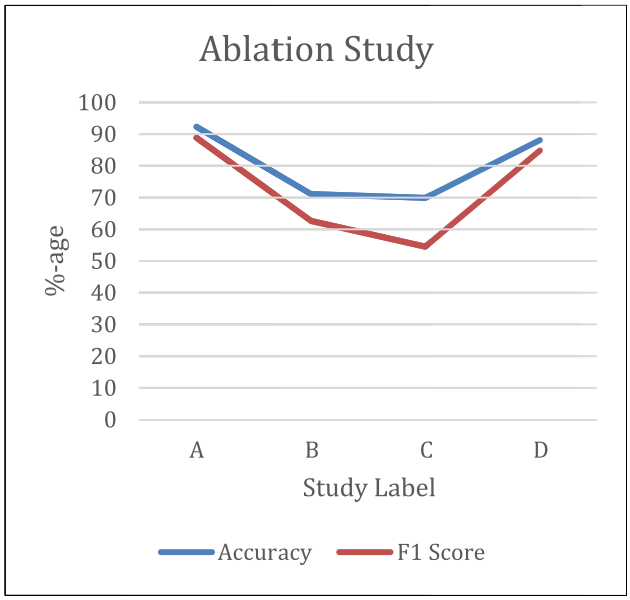

FIGURE 13. Accuracy rate showing the effect of varying features in ablation study.

The same has been validated using precision, recall and F1score on test set images from the OASIS dataset.

The results show that the Set A achieves the best result in terms of precision, and accuracy, accordingly, which is the same that is showing as per our proposed model achieving the accuracy of 92.39%. The results on recall, precision are shown in Table 5.

TABLE 5. Recall, F1 score and accuracy as per the ablation study.

Study Label	Accuracy	Recall	F1 Score
A	92.3	82.3	88.81
B	71.1	23.2	62.63
C	69.9	34.1	54.52
D	88.1	79.2	84.82

It makes sense as the features mentioned in the OASIS dataset when evaluated by the proposed MCapNet has the highest accuracy in comparison to the other deep learning methods strategies. An interesting finding here is that the Low recall of study B & C, show that our dataset is highly dependent on the Age and Gender feature. Low recall shows that actual positive predictions by the model at B and C is very less.

V. CONCLUSION AND FUTURE SCOPE

In this study, we have investigated the longitudinal MRI images consisting of neuroimaging test images in the OASIS dataset. Due to the hierarchical nature of data points having dimension as 373∗15, all the features have been considered for data modeling and predicting the disease. The EDA conducted on the dataset reveals that the age is the most dependent feature for Dementia. The proposed model uses the modified capsule network technique for classifying the dementia groups into demented and non-demented category. The model was implemented using MCapNet in python and results are validated by comparing with the other state-of-art machine learning models. The accuracy of 92.39% was given with the proposed technique which was the highest as compared to other techniques as shown in the results section. The proposed model can handle hierarchical modeling problems and is computationally more efficient with better accuracy as compared with the other models that used deep neural networks, CNN and original capsule network. The variations made in the original capsule network include additional feature extraction layer, computing matrix decomposition stated as transformation matrix using SVD, and calculating it not for every capsule, rather only one time in order to propagate across the network. These additions have made the model faster in terms of computation time as illustrated in ROC curve and run time graph., simple and more accurate. The same has been validated and shown in the ablation study by considering the hierarchical factor on all features in the OASIS dataset. It has been analyzed that gender and age are the two most prominent features in diagnosing the disease. Predictions were assessed on an independent validation set using several deep learning classifiers taken as benchmarks and comparing them with the proposed model. Since, implementing capsule network uses multiple steps, therefore applying capsule network on the heterogeneous data will make it computationally more expensive. This is observed as the limitation of our model, when implemented on other datasets with heterogeneity factor involved in their data. The model can also be improved by considering optimization and applying more validation on external data cohorts. The future scope of the paper can be directed to test the model on several other datasets names as ADNI, dementia bank dataset. The hospital study can also be validated as a future work using the proposed model where image processing will be a major task for verifying the accuracy of the proposed algorithm.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

APPENDIX A

See Figure 14 and 15.

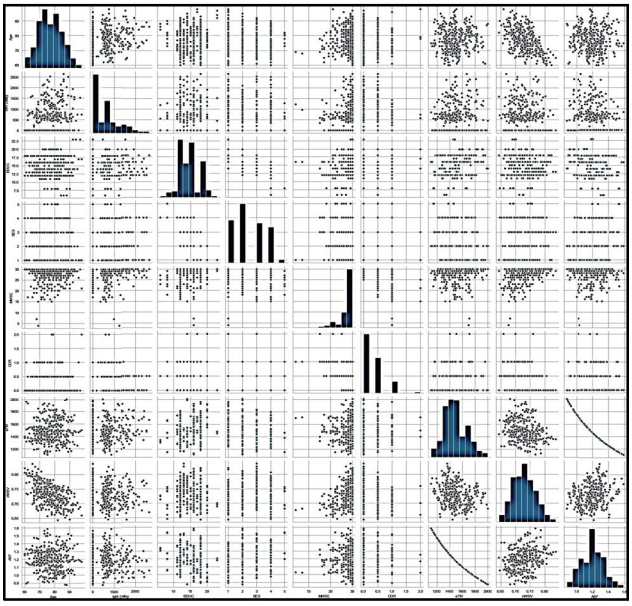

FIGURE 14. Showing Correlation between various features.

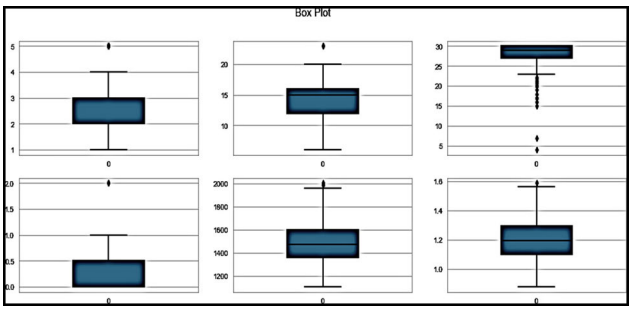

FIGURE 15. Box plot.

REFERENCES

[1] L. G. Apostolova, "Alzheimer disease," *Continuum Lifelong Learn. Neurol.*, vol. 22, no. 2, pp. 400–419, 2016.

[2] J. Lehrner, S. Kogler, C. Lamm, D. Moser, S. Klug, G. Pusswald, P. Dal-Bianco, W. Pirker, and E. Auff, "Awareness of memory deficits in subjective cognitive decline, mild cognitive impairment, Alzheimer's disease and Parkinson's disease," *Int. Psychogeriatrics*, vol. 27, no. 3, p. 357, 2015.

[3] J. Lehrner, S. Kogler, C. Lamm, D. Moser, S. Klug, G. Pusswald, P. Dal-Bianco, W. Pirker, and E. Auff, "Awareness of memory deficits in subjective cognitive decline, mild cognitive impairment, Alzheimer's disease and Parkinson's disease," *Int. Psychogeriatrics*, vol. 27, no. 3, p. 357, 2015.

[4] H. Ji, Z. Liu, W. Yan, and R. Klette, "Early diagnosis of Alzheimer's disease using deep learning," in *Proc. 2nd Int. Conf. Control Comput. Vis.*, 2019, pp. 87–91.

[5] M. A. Ebrahimighahnavieh, S. Luo, and R. Chiong, "Deep learning to detect Alzheimer's disease from neuroimaging: A systematic literature review," *Comput. Methods Programs Biomed.*, vol. 187, Apr. 2020, Art. no. 105242.

[6] R. C. Petersen, "Mild cognitive impairment," *Continuum Lifelong Learn. Neurol.*, vol. 22, no. 2, p. 404, 2016.

[7] S. Ishiai, Y. Koyama, K. Seki, S. Orimo, N. Sodeyama, E. Ozawa, E. Y. Lee, M. Takahashi, S. Watabiki, R. Okiyama, T. Ohtake, and M. Hiroki, "Unilateral spatial neglect in AD: Significance of line bisection performance," *Neurology*, vol. 55, no. 3, pp. 364–370, Aug. 2000.

[8] M. Tanveer, B. Richhariya, R. U. Khan, A. H. Rashid, P. Khanna, M. Prasad, and C. T. Lin, "Machine learning techniques for the diagnosis of Alzheimer's disease: A review," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 16, no. 1s, pp. 1–35, 2020.

[9] L. D. Deepthi, D. Shanthi, and M. Buvana, "An intelligent Alzheimer's disease prediction using convolutional neural network (CNN)," *Int. J. Adv. Res. Eng. Technol.*, vol. 11, no. 4, pp. 12–22, 2020.

[10] M. Nguyen, T. He, L. An, D. C. Alexander, J. Feng, and B. T. T. Yeo, "Predicting Alzheimer's disease progression using deep recurrent neural networks," *NeuroImage*, vol. 222, Nov. 2020, Art. no. 117203.

[11] T. Jo, K. Nho, and A. J. Saykin, "Deep learning in Alzheimer's disease: Diagnostic classification and prognostic prediction using neuroimaging data," *Frontiers Aging Neurosci.*, vol. 11, p. 220, Aug. 2019.

[12] I. Beheshti, H. Demirel, and H. Matsuda, "Classification of Alzheimer's disease and prediction of mild cognitive impairment-to-Alzheimer's conversion from structural magnetic resource imaging using feature ranking and a genetic algorithm," *Comput. Biol. Med.*, vol. 83, pp. 109–119, Apr. 2017.

[13] V. S. Nori, C. A. Hane, Y. Sun, W. H. Crown, and P. A. Bleicher, "Deep neural network models for identifying incident dementia using claims and EHR datasets," *PLoS ONE*, vol. 15, no. 9, 2020, Art. no. e0236400.

[14] J. Wang, M. J. Knol, A. Tiulpin, F. Dubost, M. de Bruijne, M. W. Vernooij, H. H. H. Adams, M. A. Ikram, W. J. Niessen, and G. V. Roshchupkin, "Gray matter age prediction as a biomarker for risk of dementia," *Proc. Nat. Acad. Sci. USA*, vol. 116, no. 42, pp. 21213–21218, Oct. 2019.

[15] H. Li and Y. Fan, "Early prediction of Alzheimer's disease dementia based on baseline hippocampal MRI and 1-year follow-up cognitive measures using deep recurrent neural networks," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 368–371.

[16] H. Li, M. Habes, D. A. Wolk, and Y. Fan, "A deep learning model for early prediction of Alzheimer's disease dementia based on hippocampal magnetic resonance imaging data," *Alzheimer's Dementia*, vol. 15, no. 8, pp. 1059–1070, Aug. 2019.

[17] M. Ganguli, Y. Jia, T. F. Hughes, B. E. Snitz, C.-C.-H. Chang, S. B. Berman, K. J. Sullivan, and M. I. Kamboh, "Mild cognitive impairment that does not progress to dementia: A population-based study," *J. Amer. Geriatrics Soc.*, vol. 67, no. 2, pp. 232–238, Feb. 2019.

[18] S. Spasov, L. Passamonti, A. Duggento, P. Liò, and N. Toschi, "A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease," *NeuroImage*, vol. 189, pp. 276–287, Apr. 2019.

[19] M. N. I. Qureshi, S. Ryu, J. Song, K. H. Lee, and B. Lee, "Evaluation of functional decline in Alzheimer's dementia using 3D deep learning and group ICA for rs-fMRI measurements," *Frontiers Aging Neurosci.*, vol. 11, p. 8, Feb. 2019.

[20] Y. Ding, J. H. Sohn, M. G. Kawczynski, H. Trivedi, R. Harnish, N. W. Jenkins, D. Lituiev, T. P. Copeland, M. S. Aboian, C. M. Aparici, S. C. Behr, R. R. Flavell, S.-Y. Huang, K. A. Zalocusky, L. Nardo, Y. Seo, R. A. Hawkins, M. H. Pampaloni, D. Hadley, and B. L. Franc, "A deep learning model to predict a diagnosis of alzheimer disease by using 18F-FDG PET of the brain," *Radiology*, vol. 290, no. 2, pp. 456–464, Feb. 2019.

[21] L. Wang, L. Sha, J. R. Lakin, J. Bynum, D. W. Bates, P. Hong, and L. Zhou, "Development and validation of a deep learning algorithm for mortality prediction in selecting patients with dementia for earlier palliative care interventions," *JAMA Netw. Open*, vol. 2, no. 7, 2019, Art. no. e196972.

[22] A. Ortiz, J. Munilla, J. M. Górriz, and J. Ramírez, "Ensembles of deep learning architectures for the early diagnosis of the Alzheimer's disease," *Int. J. Neural Syst.*, vol. 26, no. 7, Nov. 2016, Art. no. 1650025.

[23] S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, and D. Feng, "Early diagnosis of Alzheimer's disease with deep learning," in *Proc. IEEE 11th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2014, pp. 1015–1018.

[24] C. Park, J. Ha, and S. Park, "Prediction of Alzheimer's disease based on deep neural network by integrating gene expression and DNA methylation dataset," *Expert Syst. Appl.*, vol. 140, Feb. 2020, Art. no. 112873.

[25] S. Karlekar, T. Niu, and M. Bansal, "Detecting linguistic characteristics of Alzheimer's dementia by interpreting neural models," 2018, *arXiv:1804.06440*. [Online]. Available: http://arxiv.org/abs/1804.06440

[26] S. Risacher, A. Saykin, J. Wes, L. Shen, H. Firpi, and B. McDonald, "Baseline MRI predictors of conversion from MCI to probable AD in the ADNI cohort," *Current Alzheimer Res.*, vol. 6, no. 4, pp. 347–361, Aug. 2009.

[27] K. R. Kruthika, Rajeswari, and H. D. Maheshappa, "CBIR system using capsule networks and 3D CNN for Alzheimer's disease diagnosis," *Informat. Med. Unlocked*, vol. 14, pp. 59–68, Jan. 2019, doi: 10.1016/j.imu.2018.12.001.

[28] R. Aly, S. Remus, and C. Biemann, "Hierarchical multi-label classification of text with capsule networks," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics Student Res. Workshop*, 2019, pp. 323–330.

[29] I. Izonin, R. Tkachenko, N. Kryvinska, and P. Tkachenko, "Multiple linear regression based on coefficients identification using non-iterative SGTM neural-like structure," in *Proc. Int. Work-Conf. Artif. Neural Netw.* Cham, Switzerland: Springer, 2019, pp. 467–479.

[30] N. Melnykova, N. Shakhovska, M. Gregus, V. Melnykov, M. Zakharchuk, and O. Vovk, "Data-driven analytics for personalized medical decision making," *Mathematics*, vol. 8, no. 8, p. 1211, Jul. 2020.

[31] R. Aly, S. Remus, and C. Biemann, "Hierarchical multi-label classification of text with capsule networks," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics Student Res. Workshop*, 2019, pp. 323–330.

[32] Z. Wang, J. Xin, Z. Wang, H. Gu, Y. Zhao, and W. Qian, "Computer-aided dementia diagnosis based on hierarchical extreme learning machine," *Cognit. Comput.*, vol. 13, no. 1, pp. 34–48, Jan. 2021.

[33] J. Islam and Y. Zhang, "Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks," *Brain Informat.*, vol. 5, no. 2, pp. 1–14, Dec. 2018.

[34] K. Dev, S. A. Khowaja, A. S. Bist, V. Saini, and S. Bhatia, "Triage of potential COVID-19 patients from chest X-ray images using hierarchical convolutional networks," *Neural Comput. Appl.*, pp. 1–16, Feb. 2021, doi: 10.1007/s00521-020-05641-9.

[35] *Oasis Dataset*. Accessed: Dec. 10, 2020. [Online]. Available: https://www.oasis-brains.org/

[36] K. Chakraborty, S. Bhatia, S. Bhattacharyya, J. Platos, R. Bag, and A. E. Hassanien, "Sentiment analysis of COVID-19 tweets by deep learning classifiers—A study to show how popularity is affecting accuracy in social media," *Appl. Soft Comput.*, vol. 97, Dec. 2020, Art. no. 106754.

[37] S. Sabour, N. Frosst, and G. E Hinton, "Dynamic routing between capsules," 2017, *arXiv:1710.09829*. [Online]. Available: http://arxiv.org/abs/1710.09829

[38] 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," *J. Cogn. Neurosci.*, vol. 22, pp. 2677–2684, 2010, doi: 10.1162/jocn.2009.21407.

[39] M. Alojail and S. Bhatia, "A novel technique for behavioral analytics using ensemble learning algorithms in E-commerce," *IEEE Access*, vol. 8, pp. 150072–150080, 2020.

[40] S. K. Srivastava, S. K. Singh, and J. S. Suri, "A healthcare text classification system and its performance evaluation: A source of better intelligence by characterizing healthcare text," in *Cognitive Informatics, Computer Modelling, and Cognitive Science*. New York, NY, USA: Academic, 2020, pp. 319–369.

[41] S. Basheer, K. K. Nagwanshi, S. Bhatia, S. Dubey, and G. R. Sinha, "FESD: An approach for biometric human footprint matching using fuzzy ensemble learning," *IEEE Access*, vol. 9, pp. 26641–26663, 2021.

SHAKILA BASHEER is currently an Assistant Professor with the Department of Information Systems, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia. She has more than ten years of teaching experience and has published more technical articles in international journals, proceedings of international conferences, and edited chapters of reputed publications. She has worked and contributed in the field of data mining, image processing, and fuzzy logic. Her research interest includes data mining algorithms using fuzzy logic. She is also working on data mining, vehicular networks, machine learning, block chain, vehicular networks, and the IoT.

SURBHI BHATIA is doctorate in computer science and engineering with Banasthali Vidyapith, India. She is also an Assistant Professor with the Department of Information Systems, College of Computer Sciences and Information Technology, King Faisal University, Saudi Arabia. She has rich eight years of teaching and academic experience. She is in the Editorial Board Member with Inderscience Publishers in the International Journal of Hybrid Intelligence and also in several IEEE conferences. She has published seven national and international patents. She has published more than 30 articles in reputed journals and conferences in high indexing databases. She has delivered talks as a keynote speaker in IEEE conferences and in faculty development programs. She has successfully authored and edited books from Springer and Elsevier. She has been an active researcher in the field of data mining, machine learning, and information retrieval.

SAPIAH BINTI SAKRI received the B.Sc. degree in computer science from Universiti Kebangsaan Malaysia, Malaysia, in 1985, the M.Sc. degree in information science from the University of Malaya, Malaysia, in 1997, and the Ph.D. degree in information security from Universiti Kebangsaan Malaysia, in 2007. Her Ph.D. thesis was on information security management. She was with the Prime Minister's Department, Malaysia, for 30 years. She was the Director of the Information Communication Technology Policy and Strategic Division, Prime Minister's Department. She is currently an Assistant Professor with the Department of Information Systems, Princess Nourah Bint Abdulrahman University, Saudi Arabia. Her research interests include information security management, cloud computing security, data mining, machine learning, deep learning, and data science.