Predicting the Relation between Education Level, Age and Alzheimer's Disease by Simple Linear Regression

Samah Abuayeid, Hosam Alhakami, Abdullah Baz, Tahani Alsubait
s44180807@st.uqu.edu.sa, hhhakam@uqu.edu.sa, aobaz01@uqu.edu.sa, tmsubait@uqu.edu.sa
College of Computer and Information Systems, Umm Al-Qura University, Saudi Arabia

Summary
Alzheimer's disease (AD) is a progressive perturbation of brain cells that could cause decay in human social behavior. Moreover, AD is considered the most common reason for dementia that affects many people in old age. There is a demand for understanding AD risk factors to decrease the possibility of AD. These factors are classified into two categories: modified factors (lifestyle and education) and unmodified factors (age, gender, and genetic). Data analytics and machine learning techniques have been introduced in bioinformatics research to diagnose, predict, and prevent AD. The primary aim of this research is to study the relationship between education level and AD, where education level is considered as a modified risk factor. Additionally, to improve our results, we also study the relationship between AD and age as an unmodified risk factor, and then we compare the two sets of results. We have used a publicly available dataset from the Kaggle community. We have built two linear regression models where the first studies the relation between AD and education level and the second studies the relationship between patients' age and AD. We have observed that education level affects AD patients negatively if only multiple risk factors are available in the patient's environment, and that will increase the possibility of AD conversion to dementia. In general, the results demonstrate the ability of linear regression in predicting that the combination of AD risk factors could affect AD people negatively. These results encourage taking measures to enhance the AD patient's environment and reduce the number of risk factors.

Key words:
Alzheimer's disease; Dementia; MCI; False memory; Machine learning; data analysis; Linear regression.

1. Introduction

Recently, Alzheimer's disease became the most common cause of dementia for many people at old ages. According to a study presented by the world Alzheimer report 2018 [1], approximately 50 million people in the world had been affected by Alzheimer's disease in 2018 and this number is expected to be triple by 2050. Yet, many clinical studies proved that the symptoms of Alzheimer usually become visible after the age of 60. However, other researchers believe that some forms of Alzheimer's could appear very early at the age of 30 to 50. Additionally, AD, in some cases, becomes hard to predict until the patient reaches a moderate level of AD. In some stages of AD, the patients suffer from different kinds of false memory problems and lack of awareness. Awareness defined as the ability to hold a meaningful or realistic perception or appraisal response to an environment. In a study by O'Shaughnessy et al. [2], the researchers reported that 12.5% of AD patients are in the level of dementia where they had awareness problems. Additionally, awareness problems could lead to severe cognitive and functional deficits in which the patients cannot live independently and need full-time care. A related perspective is what is referred to as false memory which can be illustrated in the following use case: "Assume you see a person in the street, and you believe you've met him someplace, but you're not certain where. Fairly quickly, you discover you've hadn't met him before". Generally speaking, in our everyday life, we may experience the false memory situation. El Haj et al. [3] have discussed the consequences of false memory on AD patients and found that it could lead the patient to act upon his false belief. They also discuss different types of false memory in AD, such as false item memory, false context memory, and false autobiographical memory. The previous problems and many others have been the concern of scientific research to reduce AD symptoms. Along similar lines, Hassan et al. [4] studied the relationship between yoga practice, and its effects on AD. They argue that the biological stamp of AD includes the presence of plaques as well as neurofibrillary tangles containing a protein called tau. Moreover, these proteins damage the normal function of neurons and cause AD. In addition, they reported that the neurotransmitters excrete during yoga practices could reverse the effects of AD. Moreover, they choose to drive the model from human-induced stem cells because it could generate new neural tissues, which could prevent AD. On the other hand, Lancaster et al. [5] have developed a smartphone gallery game within the Mezurio smartphone application, which is an application for clinical research studies. The gallery game targets the episodic memory by giving a group of daily tasks for each participant, each connected with a one recognition memory and recall test. In their study [5], an experiment starts with 35 healthy adults between 40 to 59 years old who have a possible risk of AD and dementia. The authors have reported that the smartphone gallery game had performed a reasonable assessment of episodic memory. Besides, machine learning, deep learning, and data analysis had proved their benefit in early prediction AD or reducing its symptoms by analyzing AD different affecting factors. In this paper, we work on the relation between AD and level

Manuscript received April 5, 2020
Manuscript revised April 20, 2020

of education associated with age with an aim to investigate the possibility of reducing the AD conversion to dementia. We generally report that enhancing AD people's education could lead to improving awareness and reducing false memory problem.

The paper is structured as follows: Section 2 presents an overview of AD. Section3 consists of reviews of the related work. Section 4 presents the materials and methods, which is divided into 4 subsections. Section 5 introduces the proposed model. The last section concludes the paper.

2. An Overview of Alzheimer's Disease

AD addressed considerable changes in the behavioral and structural functions of the human brain. However, AD can be defined as a complex multi-factor brain degenerative disorder disease. In this section, we briefly discuss some critical information about AD. Firstly, the main symptom of AD is memory loss, which differs from ordinary people's memory lapses. In particular, in the AD memory loss situation, a patient could suffer from remembering recent actions or conversations. Additionally, people with AD could repeat statements and questions multiple times, get lost in their familiar places, and have difficulty in recalling the right words to identify an object. Besides, they could have trouble in abstract concepts thinking such as numbers also; they cannot make decisions or judgments in everyday situations. On the other hand, AD risk factors are classified into two categories: modified or unmodified. Modified risk factors include the factors that can be controlled and changed to reduce the possibility of developing Dementia. These factors include lifestyle, environment, education, head injuries, and social isolation. In comparison, unmodified risk factors include factors that cannot be controlled such as age, gender and genetics. Moreover, Dementia is defined as a clinical syndrome because many kinds of diseases can cause it, but Alzheimer's is the most common one. Many people have memory problems by aging; however, Dementia is different because it affects human ability to perform everyday tasks. Moreover, it can affect the ability of self-feeding. Dementia patients could not only forget the place of their things; but they may even forget the meaning of words. Additionally, the most known types of Dementia are vascular Dementia and mixed Dementia. Many people have conflicted between Dementia and AD. In 2019, a researcher [6] has proved that there is some similarity in AD and vascular dementia symptoms, which makes it difficult to tell if a person has AD or Vascular Dementia. However, understanding these differences could help in estimating the risk factors and possible treatment of the two diseases. On the other hand, the relation between Mild Cognitive Impairment (MCI) and AD is that some MCI patient's situations could convert to AD. Moreover, MCI is considered as part of the normal aging situation, but AD is not. Furthermore, MCI has been defined as a pathological level between normal aging and Dementia, which could be caused by AD. Additionally, there are many classifications for AD levels and the most straightforward classifications are shown in Fig 1.

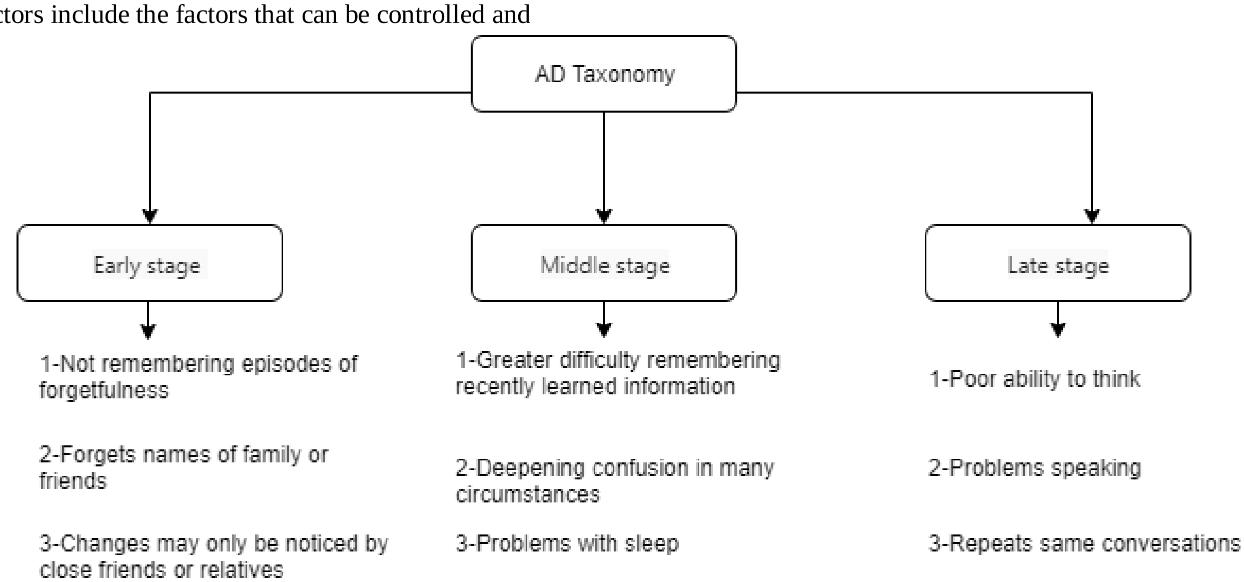

Fig. 1 Alzheimer's disease Taxonomy.

3. Related Work

This section presents a comparative review of current studies in the research area. Depending on the methods and techniques covered in the reviewed studies, we have divided this section into two main subsections: machine learning investigations and deep learning investigations.

3.1 Deep Learning Investigations

Basaia et al. [7] reported that machine learning methods are time-consuming because it needs complex images processing comparing to deep learning methods. As a result, they preferred to build a deep learning algorithm to reduce the effort required in feature extraction and image processing. Another research in the same area presented by Gopan et al. [8] had proposed a model for recognizing AD brain image to understand how much brain regions has affected by AD. The experiment follows seven main steps starting with selecting an image and converting it to a gray-scale image then filtering with a median filter to remove noise. Next, they implemented image segmentation using K-means as a clustering algorithm. K-means is an unsupervised clustering algorithm that classifies the data depending on their distances. Finally, the extracted image is compared with a training dataset using a support vector machine (SVM). Moreover, they had to repeat the comparison more than one-time using random forest (RF) and bagged SVM. As a result, the authors reported that the accuracy of bagged SVM is the best. Moreover, clinical investigations in predicting the rate and the possibility of conversion patient situation from MCI to AD and finally dementia need more biological markers with a variety of roles and functions implemented on complex combination MRI and PET images, which could be time-consuming [20]. A recent study by Franzmeier et al. [9] has presented a model to predict the rates of cognitive decline in AD by using support vector regression as a machine learning method to AD bio-markers resulted from MRI and Positron Emission Tomography (PET). The proposed model has given an accurate result with the least possible time compared with typical clinical prediction. On the other hand, Huber-Carol et al. [10] have applied a comparison between stochastic model, logistics regression, and neural network in the accuracy of predicting AD patient situation for the next four years using the R language. The researchers have used 17 risk factors varied between modified and unmodified risk factors. Additionally, they used a dataset consisting of 5003 patients information offered by a hospital in Paris. As a result, they realized that the education level does not affect the AD patient's situation negatively as other studies assumed. Moreover, Huber-Carol et al. [10] have proved that logistics regression and neural network give close results, although that neural network is the commonly used machine learning method with big data analysis. Park et al. [11] have used neural networks to build an AD prediction model. The researchers had adopted the idea of using deep learning methods rather than machine learning aiming to improve the result of AD prediction as it was also reported in [10]. The model in [11] has examined two different datasets, gene expression and DNA methylation, which consist of different characteristics. The verity of datasets' aspects enforced the researches to use multiple steps of features selection and then implement the neural network on the datasets. As a result of the experimentation, the following points have been reported:

- Using two datasets will improve prediction accuracy.
- Using deep learning rather than machine learning improves AD prediction performance.

Moreover, Islam and Zhang [12] designed a convolution network for AD detection and classification using MRI images. The proposed convolution network builds on the work introduced by Szegedy et al. [21]. It consists of seven 3x3 layers, one 1x1 layer, and a pooling layer. Islam and Zhang have used the OASIS dataset offered by Howard Hughes Medical Institute at Harvard University. The experiment had achieved 73.75% AD prediction accuracy. Neural network and feature selection enhanced AD prediction, as proved by a study by Razavi et al. [13]. The proposed model in [13] consists of a two-level learning method for diagnosing AD. At the first level, they used the sparse filtering for brain images features extraction, and the learned features obtained, wherein the second level SoftMax regression had been applied on the learned features to classifying patient health status. Additionally, the model had been applied on the ADNI dataset, and the model performance measured using the F-score measurement. As a result, Razavi et al. assumed that unsupervised learning methods could improve AD diagnosing, Table 1 summarizes deep learning studies presented in this subsection.

Table 1: Compression between deep learning studies

Paper	*Goal*	*Method*
[7]	Predict AD	Deep learning
[8]	Brain affected rate	K-mean, SVM
[9]	Predict cognitive decline	SRV
[10]	Next 4 years situation	Neural network and logistics regression
[11]	AD Prediction	Neural network
[12]	AD Prediction	convolution network
[13]	AD diagnosing	Neural network

3.2 Machine Learning Investigations

Bin-Hezam and Ward [14] have implemented machine learning algorithms on dementia modified risk for disease early diagnoses. The study classified the diagnosis results into three categories: cognitively normal, mild cognitive impairment, or dementia. Another research in the same area

[15] had proposed a machine learning model for early prediction of dementia using a dataset offered by dementia center in Korea. So et al. [15] had used multiple machine learning algorithms such as Naive Baye, support vector machine (SVM), and random forest for performance testing. Grassi et al. [16] used machine learning algorithms to predict AD conversion to dementia. Along similar lines, Tanveer et al. [17] have reviewed some machine learning studies on AD and reported that many machine learning AD research had used SVM and Kernel function for features selection and extraction of the brains or genetics images to diagnose AD earlier. Additionally, Lourida et al. [18] presented a study on dementia and discovered that there is a strong relationship between the increased risk of dementia and people's lifestyle, in which a healthy lifestyle could decrease the risk of dementia and vice versa. According to many studies, AD and dementia can affect people at the age of 30 to 50, and this type of AD is considered an early AD compared with the one that affects older people. Hall et al. [19] have used data analysis and machine learning methods to predict dementia risk at different ages. The study has proposed that the early prediction of dementia could be easier than a late prediction that involves older people. On the other hand, Basaia et al. [7] used deep learning algorithms to predicate AD in normal people or AD development in MCI patients. The study used the cross-sectional brain structural Magnetic Resonance Imaging (MRI) scan founded in the ADNI dataset. Noticeably, many research studies have used SVM as a machine learning algorithm in AD prediction.

4. Methods and materials

4.1 Research Methodology

The research presented in this paper is inspired by the work presented by Kologlu et al. [22], which uses linear regression to predicate football players' market value using players' physical and performance factors. The footballers have been chosen from Europe leagues between season 2017 and 2018 in the age from 20 to 33. Along similar lines, in a different context, our primary aim of this research is to predict the relationship between education level and AD, where education level is considered as a modified risk factor for AD. Additionally, to improve our result, we choose one unmodified risk factor, which is Age. Additionally, we compare our results to the results reported in [10], in which the authors report that the level of education does not negatively affect AD patients. However, the chosen risk factors must be a combination of modified and unmodified factors, so we decided to select age as an unmodified factor and education as the main modified risk factor. Additionally, our proposed model depending on (Group) as another essential attribute from the dataset, which will be discuss in more detail in the next subsections. By using this information, we will be able to implement two simple linear regression models to perform a precise prediction for the relation between AD and education. Moreover, we will compare the result with the relationship between an unmodified risk factor (in our case, Age) and AD. The experiments conducted for the research was performed using Jupyter Lab 1.2.6 as a programming environment, Python 3 as a programming language, and Scikit-learn package for machine learning and model training.

4.2 Dataset

For the lack of available open-source AD datasets, we prefer to use the Longitudinal MRI Data in Non-demented and Demented Older Adults, which was offered by the Kaggle community. The dataset has been used in many Kaggle AD projects, and it consists of different data for 150 subjects aged from 60 to 96. Each subject from the 150 subjects was investigated on two or more clinical visits, separated by one year or more for a total of 373 sessions. The subjects include men and women, and 72 of the subjects are classified as Non-demented, where 64 of them classified are as Demented at their initial clinical visits. Additionally, a minority number of subjects was classified as Converted. Table 2 shows the dataset abbreviation.

Table 2: Longitudinal MRI Dataset

Abbreviation	***Description***
Subject ID	ID assigned to the subject
MRI ID	diagnostic exam ID
Group	demented or non-demented
Visit	clinical visits number
MR Delay	Sessions delay time
M / F	male or female
Hand	left or right handed
Age	subject age
EDUC	subject years of education
SES	subject's Socioeconomic Status
MMSE	Mini Mental State Examination
CDR	Clinical Dementia Rating
ETIV	Estimated total intracranial volume
nWBV	Normalize Whole Brain Volume

4.3 Data Exploration and Preprocessing

In this section, we discuss the data exploration step, which is the primary data investigation which helps in better understanding the used dataset. Data exploration could give a proper perspective for the techniques and methods suitable to achieve research goals and enhance the resulting quality. To implement linear regression on the chosen attributes from the dataset we need to ensure that all these attributes have the same data type; In other words, we need to replace the nonnumeric values with corresponding numeric values. Moreover, Fig 2 shows that the Group attribute has a different data type than the other attributes. However, we decide to propose a new numeric value for each value stored in the Group attribute. To illustrate this, we replace Non-demented with 1, Demented with 2, and Converted with 3. Then, we extracted the essential attributes from the dataset, which are (Group, Age, EDUC). The target risk factor is EDUC (education level), and the main predictor risk factor is Group which is divided into three categories Demented (there is a possibility to convert AD to dementia), Non-demented (there is no possibility to convert AD to dementia), and Converted (already converted to dementia).

	Subject ID	MRI ID	Group	Visit	MR Delay	M/F	Hand	Age	\
0	OAS2_0001	OAS2_0001_MR1	Nondemented	1	0	M	R	87	
1	OAS2_0001	OAS2_0001_MR2	Nondemented	2	457	M	R	88	
2	OAS2_0002	OAS2_0002_MR1	Demented	1	0	M	R	75	
3	OAS2_0002	OAS2_0002_MR2	Demented	2	560	M	R	76	
4	OAS2_0002	OAS2_0002_MR3	Demented	3	1895	M	R	80	
..	...	...	...	...	...	..	...	...	
368	OAS2_0185	OAS2_0185_MR2	Demented	2	842	M	R	82	
369	OAS2_0185	OAS2_0185_MR3	Demented	3	2297	M	R	86	
370	OAS2_0186	OAS2_0186_MR1	Nondemented	1	0	F	R	61	
371	OAS2_0186	OAS2_0186_MR2	Nondemented	2	763	F	R	63	
372	OAS2_0186	OAS2_0186_MR3	Nondemented	3	1608	F	R	65	

Fig. 2 Non-numeric values.

We also extract Age as unmodified target risk factor. Consequently, we used multiple risk factors, and compare two simple linear regression models in terms of their accuracy. After that, we explored the missing values in the previously mentioned attributes by counting the missing values if it occurs in each attribute, and discard them to enhance the final results; Table 3 shows some details.

Table 3: Stats after pre-processing step

Number of instances	373
Number of attributes	3
Number of missing values	0
Number of rows in original data	373
Number of rows after discarding missing values	373

To be more accurate and to ensure that the Group attribute had changed to numeric, we try to explore the outlier, which is data instances with different data types from the other attributes in the dataset shown in Figure 3.

Table 3: Missing values exploration

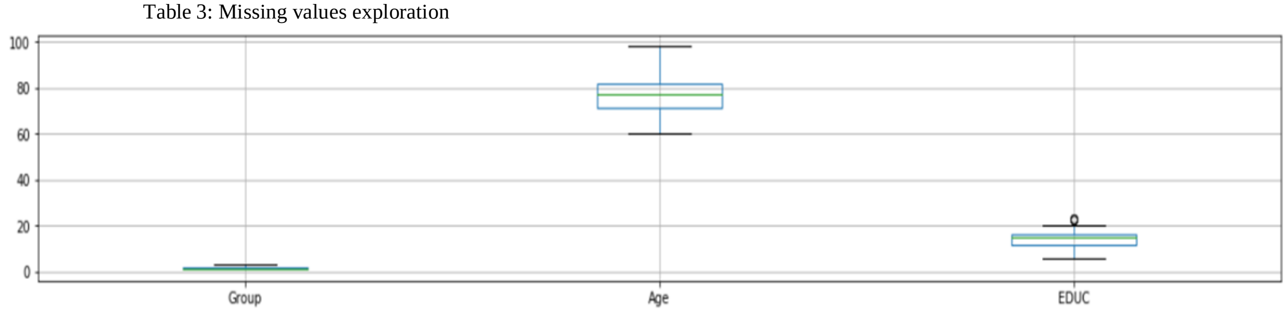

Fig. 3 Outliers.

Next, we implement a data sampling approach to facilitate data reduction. However, it's precise form Figure 4 that Group attribute values had changed from the previous steps. Additionally, using the discretization method for categorizing the data in some of the attributes will get a better understanding of the data, which is shown in Figure 4.

4.4 Linear Regression

Linear regression is a machine learning statistical analysis technique used in data analysis research to find the linear relation between factors (i.e., variables). Additionally, the factors are divided into a target (response) factor and the independent (predictor) factor. The linear regression algorithm defines the best value for the intercept and the line slope, which is given in equation 1:

$$Y = mX + b \tag{1}$$

where Y, X are variables, m is line slope and b is the intercept. There are two types of linear regression:

- Simple linear regression which studies the relation between one target(independent) factor and its corresponding predictor (dependent) factor.
- Multiple linear regression which studies the relation between two or more target (independent) factors and their corresponding predictor (dependent) factors.

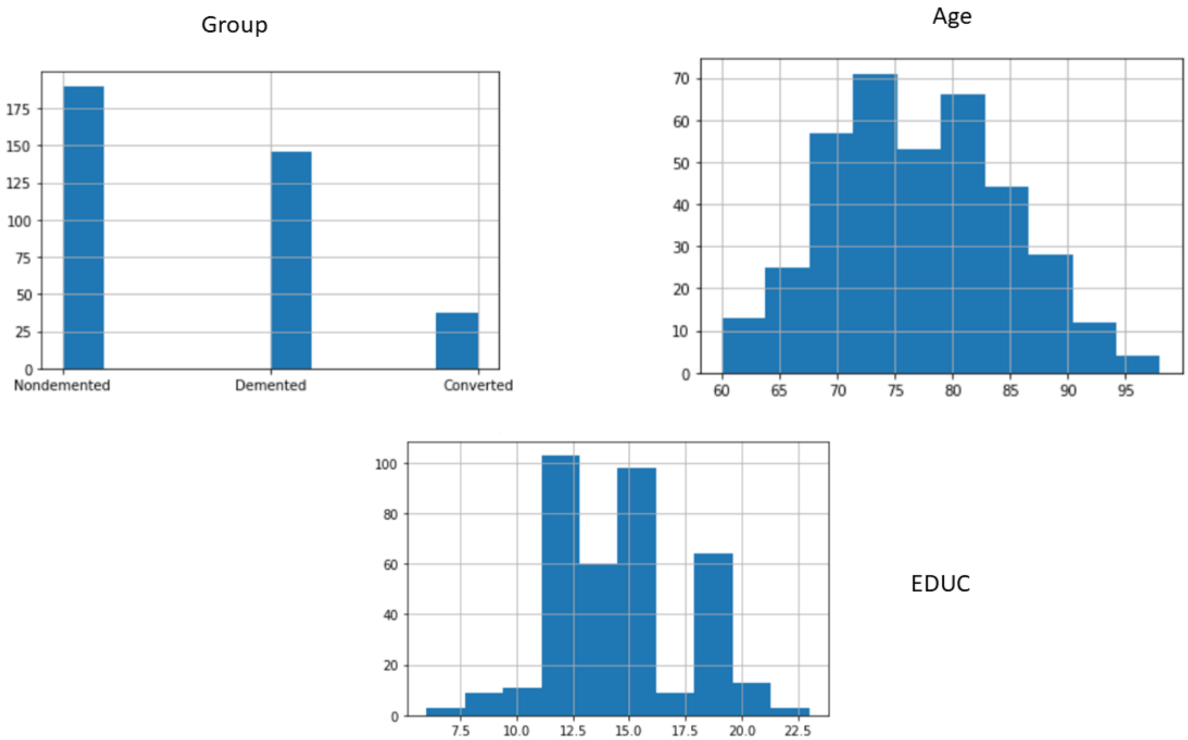

Fig. 4 Data discretization results.

5. The Proposed Model

We have implemented two different simple linear regression methods rather than multiple linear regressions because of the limited number of our factors and the accuracy of our prediction model. Moreover, the relation between EDUC and Group attribute will not be accurate if we add the Age attribute as the second prediction factor in multiple linear regression. We propose implementing two separate linear regression models, one for predicting the relation between EDUC and Group and another one for predicting the relationship between Age and Group and then comparing the results from the two models. In the first step, we propose the predictor variable Y and the target variable X as in Equation 1, so for the education prediction regression model, the target= EDUC and predictor= Group. Moreover, in the Age prediction model target= Age and predictor= Group. Then we divided the extracted dataset into the training dataset and the test dataset. The division step should follow some constraints, which is 20% for test and 80% to train the model.

5.1 Model Performance Evaluation

The primary purpose of this section is to ensure that the model addressed the minimum rate of the errors. We start by calculating Y-predict from X-test using the linear regression class in Python3 as in equation 2, where “regressor” refers to an object from a linear regression class.

$$Y - pred = regressor{:}predict(X - test) \tag{2}$$

Then we study the difference between actual value (y-test) and the prediction value (y-predict). Additionally, it’s clear from Fig 5 that the prediction value becomes close to the actual value when Actual=2, but prediction value becomes far from actual value when Actual= 3 or Actual=1. In other words, education level does not affect the Non-demented AD patients (Actual=1) or the Converted AD patients (Actual=3), but it has a parallel relation with the Demented AD patients, where (Actual=2). Additionally, the rate of error varies between the Actual values. As a result, we calculate three standard metrics used to evaluate the accuracy of continues variables:

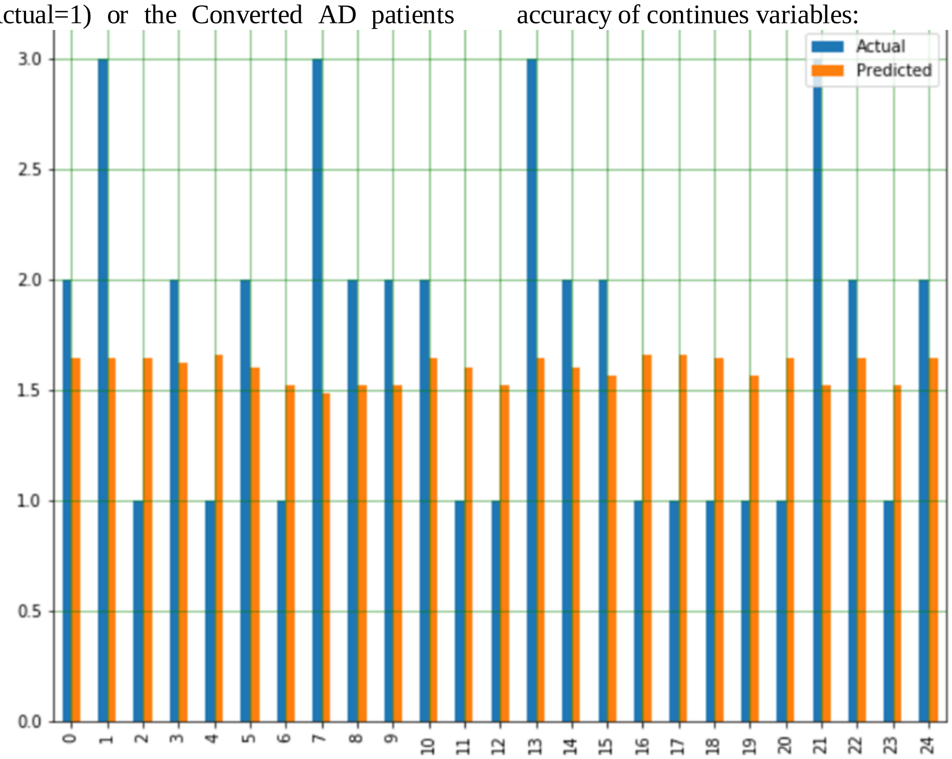

Fig. 5 Comparing model actual and prediction results.

- Mean Absolute Error (MAE): It's the most straightforward regression error metric that offers measurement of the average rate of errors in a group of predictions. In other words, it presents the average from a test sample of the absolute differences between prediction values and actual values, where all disputes have equal weight which can calculated using equation 3. Moreover, MAE will never eliminate the negative values because it always calculates the absolute function.

$$MAE = \frac{1}{n}\sum_{j=1}^{n}\left|y_j - y_j^*\right| \quad (3)$$

- Mean Squared Error (MSE): It measures the average square of errors between actual values and the prediction values, as in equation 4. However, because MSE calculates the average squared of the errors, it can give large results, for that it's difficult to compare MAE with MSE.

$$MSE = \frac{1}{n}\sum_{j=1 \to n}\left(y_j - y_j^*\right)^2 \quad (4)$$

- Root Mean Squared Error (RMSE): As MAE, the root mean squared error is a measurement of the average rate of errors in a group of predictions, but it's the square root difference between prediction values and actual values as given in equation 5.

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1 \to n}\left(y_j - y_j^*\right)^2} \quad (5)$$

After implementing the three metrics on our simple linear regression model, we observed the following results:

- Mean absolute error = 0.58.
- Mean squared error = 0.42.
- Root mean squared Error = 0.64.
- RMSE is greater than or equal MAE.
- MAE is greater than MSE.

It is hard to compare the results of MSE and MAE as we mentioned before because MSE always addresses square values, which are larger than the actual values, but in our model, MAE had a greater value than MSE with approximately 0.16 percentage. However, the results may change if we implement the model on a larger dataset. Additionally, according to the observation, the rate of errors in the training model is less than the expectation from using a small dataset to train the linear regression model. Moreover, it's clear form MAE and RMSE values that the errors have the same magnitude, where the RMSE result became approximately greater than or equal to the MAE by 0.06. Consequently, the model had a good performance with a small dataset (373 subjects) and an error proportion of less than 1.

6. Findings and Discussion

The primary aim of this research is to study the relationship between education level and AD, where education level is considered as a modified risk factor. Additionally, we compare this result to our investigation of the relationship between AD and Age as an unmodified risk factor. The last important attribute from the dataset was Group, which consists of the outcome of patients diagnose situations (Non-demented, Demented and Converted). Moreover, we have built two separate simple linear regression models. The first model studies the relation between Group attributes and education level. The second model examines the relationship between patients' age and the Group attribute. Furthermore, after predicting the relation between education and patients' situation (Group attribute) from the first model, we used the second model for comparing education risk factor effect with the effect of unmodified risk factor (age). However, Fig 6 shows two scatter plots A and B, that visualized the result from the two models. However, describing any scatter plot need identifying four main characteristics, which are form (linear or non-linear), direction (positive or negative), strength (strong, moderate, and weak) and outlier.

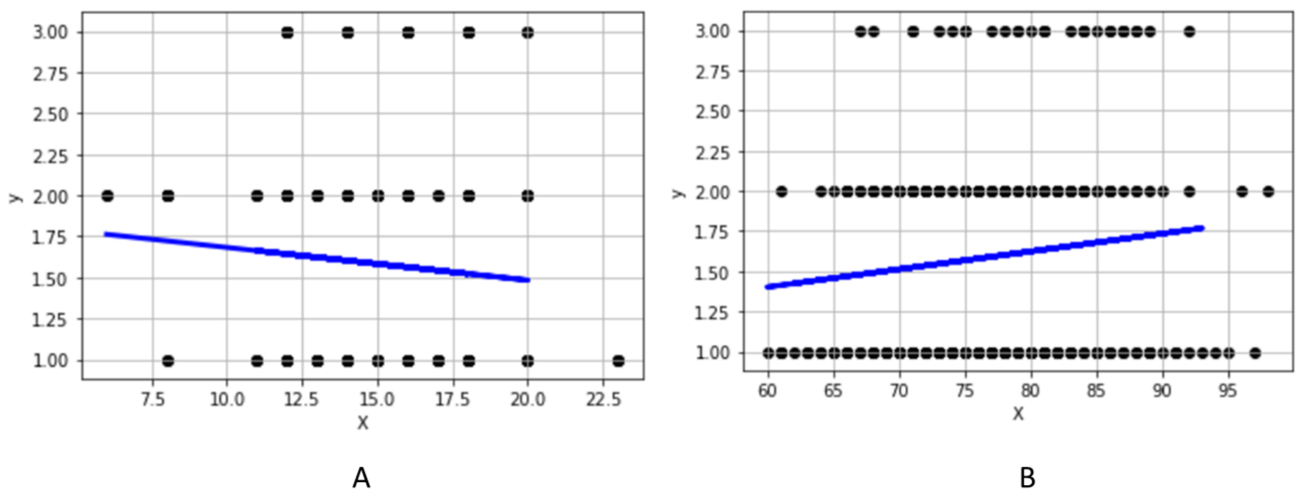

Fig. 6 (A) EDUC and Group, (B) Age and Group.

The plot in Fig 6-A presents the relation between EDUC and Group as a linear, moderate relationship with no outlier, but the direction of the relation is constant correlation almost near being negative. On the other hand, Figure 6, part B presents the relationship between Age and Group as linear, strong with no outlier, but the direction of the relation is constant correlation almost near being positive. However, we have demonstrated that the relationship between education and AD is linear, moderate, and it's not positive, which is consistent with the result reported by [10]. In other words, if education level was high, and AD patient does not address other risk factors, then education will not affect AD negatively, which means the low possibility of converting AD to dementia. To illustrate this, if AD patients follow a good lifestyle with a high rate of functional socioeconomic status, there will be no negative effect from the education level. However, if other risk factors available, then the education level will address an adverse effect on AD patients, and that could cause a conversion from AD to dementia. Moreover, we observed that even if age is one of the common unmodified risk factors with a strong relation with AD, but it's not necessarily hurting AD patients. In other words, AD and dementia are indeed common symptoms for normal aging, but that also depends on the other risk factors. Finally, we admit that the combination of education level and age as risk factors could affect the AD patient's situation negatively and cause a conversion from AD to dementia.

7. Conclusion and future work

We built two separate simple linear regression models: one for studying the relationship between education level and patient situation (Group), where the other for the relationship between patient age and patient situation (Group), then we have compared the two results. The comparison between the models had helped us in understanding in which direction does education level affects AD compared to age. Moreover, simple linear regression has achieved a reliable prediction performance with a minimum rate of errors. However, the main limitation of our model is the number of considered factors, which are two factors. Additionally, there is no way to use a combination of 3 or more risk factors in a simple linear regression model. On the other hand, we have tried to implement multiple linear regression using more than two risk factors, but the rate of prediction errors was high, and we could not identify the relationship between education and AD. This can be further investigated in future research. To conclude, this paper has admitted that education level affects AD patients negatively if only multiple risk factors are available in the patient's environment, and that will increase the possibility of conversion AD to dementia. In the future, we will generalize our model using a larger dataset, such as the ADNI dataset.

References

[1] C. Patterson, "World alzheimer report 2018: the state of the art of dementia research: new frontiers," Alzheimer's Disease International (ADI): London, UK, 2018.
[2] N. O'Shaughnessy, J. Chan, R. Bhome, P. Gallagher, H. Zhang, L. Clare, E. Sampson, P. Stone, and J. Huntley, "Awareness in severe alzheimer's disease: a systematic review," Aging & Mental Health, pp. 1–11, 2020.
[3] M. El Haj, F. Colombel, D. Kapogiannis, and K. Gallouj, "False memory in alzheimer's disease," Behavioural Neurology, vol. 2020, 2020.
[4] Hassan, M. Robinson, S. M. Willerth et al., "Determining the mechanism behind yoga's effects on preventing the symptoms of alzheimer's disease," Neural regeneration research, vol. 15, no. 2, p. 261, 2020.
[5] Lancaster, I. Koychev, J. Blane, A. Chinner, C. Chatham, K. Taylor, and C. Hinds, "Gallery game: Smartphone-based assessment of longterm memory in adults at risk of alzheimer's disease," Journal of Clinical and Experimental Neuropsychology, pp. 1–15, 2020.
[6] K. Javanshiri, M. Haglund, and E. Englund, "Hypertension and diabetes mellitus are features of vascular dementia, not alzheimer's disease," J Neurol Neurophysiol, vol. 10, p. 482, 2019.
[7] S. Basaia, F. Agosta, L. Wagner, E. Canu, G. Magnani, R. Santangelo, M. Filippi, A. D. N. Initiative et al., "Automated classification of alzheimer's disease and mild cognitive impairment using a single mri and deep neural networks," NeuroImage: Clinical, vol. 21, p. 101645, 2019.
[8] S. Gopan, P. Shruthi, M. Yokhasri, and M. A. Pandian, "Analysis of alzheimer's disease on mri image using transform and feature selection method," 2019.
[9] N. Franzmeier, N. Koutsouleris, T. Benzinger, A. Goate, C. M. Karch, A. M. Fagan, E. McDade, M. Duering, M. Dichgans, J. Levin et al., "Predicting sporadic alzheimer's disease progression via inherited alzheimer's disease-informed machine-learning," Alzheimer's & Dementia, 2020.
[10] Huber-Carol, S. Gross, and F. Vonta, "Risk analysis: Survival data analysis vs. machine learning. application to alzheimer prediction," Comptes Rendus M´ecanique, vol. 347, no. 11, pp. 817–830, 2019.
[11] 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 Systems with Applications, vol. 140, p. 112873, 2020.
[12] J. Islam and Y. Zhang, "A novel deep learning based multi-class classification method for alzheimer's disease detection using brain mri data," in International Conference on Brain Informatics. Springer, 2017, pp. 213–222.
[13] Razavi, M. J. Tarokh, and M. Alborzi, "An intelligent alzheimer's disease diagnosis method using unsupervised feature learning," Journal of Big Data, vol. 6, no. 1, p. 32, 2019.
[14] R. Bin-Hezam and T. E. Ward, "A machine learning approach towards detecting dementia based on its modifiable risk factors," International Journal of Advanced Computer Science and Applications, vol. 10, no. 8, 2019.
[15] So, D. Hooshyar, K. W. Park, and H. S. Lim, "Early diagnosis of dementia from clinical data by machine learning techniques," Applied Sciences, vol. 7, no. 7, p. 651, 2017.
[16] M. Grassi, N. Rouleaux, D. Caldirola, D. Loewenstein, K. Schruers, G. Perna, and M. Dumontier, "A novel ensemble-based machine learning algorithm to predict the conversion from mild cognitive impairment to alzheimer's disease using socio-demographic characteristics, clinical information and neuropsychological measures," Frontiers in neurology, vol. 10, p. 756, 2019.
[17] M. Tanveer, B. Richhariya, R. Khan, A. Rashid, P. Khanna, M. Prasad, and C. LIN, "Machine learning techniques for the diagnosis of alzheimer's disease: A review," ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 2019.
[18] Lourida, E. Hannon, T. J. Littlejohns, K. M. Langa, E. Hypp¨onen, E. Ku´zma, and D. J. Llewellyn, "Association of lifestyle and genetic risk with incidence of dementia," Jama, vol. 322, no. 5, pp. 430–437, 2019.
[19] Hall, T. Pekkala, T. Polvikoski, M. van Gils, M. Kivipelto, J. L¨otj¨onen, J. Mattila, M. Kero, L. Myllykangas, M. M¨akel¨a et al., "Prediction models for dementia and neuropathology in the oldest old: the vantaa 85+ cohort study," Alzheimer's research & therapy, vol. 11, no. 1, p. 11, 2019.
[20] H. Hampel, K. Broich, Y. Hoessler, and J. Pantel, "Biological markers for early detection and pharmacological treatment of alzheimer's disease," Dialogues in clinical neuroscience, vol. 11, no. 2, p. 141, 2009.
[21] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in Thirty-first AAAI conference on artificial intelligence, 2017.

[22] Y. Kologlu, H. Birinci, S. I. Kanalmaz, and B. Ozyilmaz, "A multiple linear regression approach for estimating the market value of football players in forward position," arXiv preprint arXiv:1807.01104, 2018.

Samah Abuayeid received her B.Sc. degree from King Abdulaziz university in 2009. Currently, she is a Master of Science candidate at Umm Al-Qura University, College of Computer and Information Systems, Computer Science Department.

Hosam Alhakami received his B.Sc. degree in Computer Science from King Abdulaziz University, Saudi Arabia in 2004. From 2004 to 2007, he worked in software development industry, where he implemented several systems and solutions for a national academic institution. Following that, he started his postgraduate studies in UK, where he received his MSc degree in Internet Software Systems from Birmingham University, Birmingham, UK in 2009. Then he successfully acquired his PhD in Software Engineering from De Montfort University in 2015. His research interests include algorithms, semantic web and optimization techniques. He focuses on enhancing real-world matching systems using machine learning and data analytics in a context of supporting decision-making.

Abdullah Baz received the B.Sc. degree in electrical and computer engineering from UQU, in 2002, the M.Sc. degree in electrical and computer engineering from KAU, in 2007, and the M.Sc. degree in communication and signal processing and the Ph.D. degree in computer system design from Newcastle University, in 2009 and 2014, respectively. He was a Vice-Dean, and then the Dean of the Deanship of Scientific Research with UQU, from 2014 to 2020. He is currently an Assistant Professor with the Computer Engineering Department, a Vice-Dean of DFMEA, the General Director of the Decision Support Center, and the Consultant of the University Vice Chancellor with UQU. His research interests include VLSI design, EDA/CAD tools, coding and modulation schemes, image and vision computing, computer system and architecture, and digital signal processing. Since 2015, he has been served as a Review Committee Member of the IEEE International Symposium on Circuits and Systems (ISCAS) and a member of the Technical Committee of the IEEE VLSI Systems and Applications. In 2017, IEEE has elevated him to the grade of IEEE Senior Member. He served as a Reviewer in a number of journals, including the IEEE Internet of Things, the IET Computer Vision, the Artificial Intelligence Review, and the IET Circuits, Devices and Systems.

Tahani Alsubait is a faculty member of College of Computer and Information Systems. She earned her PhD in AI and instruction from the University of Manchester. She hold a Bachelor's in Computer Science from King Saud University and a Master's from King Abdulaziz University. Her research interests include knowledge representation and reasoning, data analytics and HCI.