Mathematical Modelling of Diabetes Mellitus and Associated Risk Factors in Saudi Arabia

Entissar Almutairi, Maysam Abbod, Take Itagaki

Department of Electronic and Computer Engineering, Brunel University London, Uxbridge, UK.

Enrissar.Almutairi@brunel.ac.uk; Maysam.Abbod@brunel.ac.uk; Take.Itagaki@brunel.ac.uk

***Abstract* - Mathematical modelling has been successfully applied to the healthcare domain and epidemiological chronic diseases, including diabetes mellitus, which is classified as an epidemic due to its high rates of global prevalence. This paper models diabetes mellitus in Saudi Arabia along with the most relevant risk factors, namely smoking, obesity, and physical inactivity for adults aged ≥25 years. The aim of this study is based on developing different mathematical models for the purpose of studying the trends in incidence rates of diabetes over 15 years (1999 -2013) and to get predictions for the future level of the disease up to 2025, as this should support health policy planning and identifying the necessary costs of controlling diabetes. Different models were developed, namely Logistic Regression, Neural Networks, and Artificial Neural Networks. An overview of the performance of these models is provided to analyse their advantages and limitations. A combination of these models is performed to improve the prediction accuracy using combination methods such as AVR, WAVR, and MAJ. The combined model was validated by comparing the prediction of prevalence estimates by World Health Organization, International Diabetes Federation, and Family Health Survey from the Saudi General Authority for Statistics. Improved accuracy was achieved with this combined model in comparison to these studies.**

Keywords- mathematical modelling, diabetes, risk factors.

I. INTRODUCTION

Over recent centuries mortality rates have significantly fallen worldwide, due to improved standards of living and developments in health interventions that address the most common communicable diseases, such as cholera, typhoid, smallpox and others. Although some developing countries still face these diseases, chronic diseases associated with lifestyle factors have become the most prevalent threat to health, including cancer, cardiovascular disease, diabetes, and heart disease [1]. This shift in the pattern of diseases and their causes is called epidemiological transition, which was originally proposed by Omran in 1971 [2]. Rapidly and continuously changing lifestyles and urbanization contributed to this transition, which lead to an increase in the risk factors of non-communicable diseases. Consequently, the economic burden of these diseases is increasing around the world, particularly in developing countries [3]. Diabetes is a clear example of the phenomenon of epidemiological transition [4], which is the focus of this study. Diabetes mellitus is a significant disorder of the metabolism that leads to chronic hyperglycemia, and causes abnormalities in the metabolism of carbohydrates, fats, and proteins as a result of deficient production of insulin, resistance to insulin produced, or both [5]. The two main types of diabetes are known as type 1 and type 2. Patients with type 1 diabetes need insulin injections to survive, while type 2 diabetes, which represents most cases, is a defect in the secretion and function of insulin, meaning some diabetics of this type need insulin but most do not, as they continue to produce insulin [6].

Diabetes is a serious health problem that is growing significantly around the world because of several demographic and behavioral factors including increasing population density, urbanization, an aging population, a prevalence of obesity, and low physical activity [7]. According to the International Diabetes Federation (IDF), there were more than 460 million people with diabetes in 2019, forecast to increase to reach 578 million by 2030, and 700 million in 2045. Their published statistics indicated that there are 4 million diabetic people in Saudi Arabia (SA) [8].

This study contributes to develop different models namely, Multiple Linear Regression (MLR), Adaptive Neuro-Fuzzy Interference System (ANFIS), Artificial Neural Networks (ANN), Support Vector Regression (SVR) and Bayesian Linear Regression (BLR), in order to describe the prevalence pattern of diabetes and identify its relationship with the related risk factors.

The paper is organized as follows. Section II reviews related work. Section III presents modelling methodologies. Section IV discusses experimental methodology. Section V presents and discusses the results. Finally, section VI concludes this paper and identifies areas for future research.

II. RELATED WORK

Mathematical modelling has been successfully applied to the healthcare domain and epidemiological chronic diseases, including diabetes, and it plays an important role in the processes of description, prediction, and evaluation of these diseases, which in turn provides support and assistance to policy and decision makers [9][10]. In the last few decades, a variety of modelling studies have predicted the incidence of diabetes and its global prevalence for different countries around the world, including SA, using diverse data and methods of analysis. In [7] a mathematical model based on a set of differential equations was used to analyze estimations of diabetes in terms of occurrence, prevalence, and mortality rate. In addition, IDF [11] and another international study [12] used a logistic regression model to estimate the global prevalence of diabetes. Moreover, a further study by NCD Risk Factor Collaboration used Bayesian hierarchical model in order to make estimations for diabetes prevalence rates for different countries [13].

Each model has its own advantages and disadvantages. However, in this study, we developed different mathematical models (MLR, ANFIS, ANN, SVR, BLR) in order to take the benefits of all models and offset the individual weaknesses of each one when used individually. One of the main issues is accuracy, which can vary according to the inputs. In order to achieve uniform accuracy, these models were combined and validated in order to get an accurate and reliable predictions of prevalence rate of diabetes and the associated risk factors, with the aim of supporting health policy planning and identifying the necessary costs of controlling diabetes in SA.

III. MODELLING METHODOLOGIES

This section gives an overview of the proposed models and describes the combination methods as applied in this study.

A. Models Overview

A1. Multiple Linear Regression Model

Multiple linear regression is one of the most common types of linear regression analysis. It is an extended form of simple linear regression, with a relationship between more than two variables [14]. In predictive analysis, this technique describes the relationship between one dependent (response) variable and two or more independent (predictor) variables. The general model of multiple linear regression is:

$$Y = b_0 + b_1X_1 + b_2X_2 + \cdots + b_nX_n \tag{1}$$

Where Y is the dependent variable; $b_0, b_1, b_2, \dots\dots, b_n$ are the coefficients; and $X_1 + X_2 + \cdots + X_n$ are the independent variables.

A2. Support Vector Regression Model

Support Vector Machines (SVM) is a popular method developed by Vapnik [15]. The generalized concepts of SVM have been applied to regression problems such as modelling and prediction, and accordingly called Support Vector Regression (SVR). SVR has been effectively utilized to deal with forecasting issues in many areas as diverse as pharmacology, economics, and power systems analysis. SVR is less popular than SVM, but it has been verified that it is a valuable technique in estimating the real value of a function [16]. One of the most useful features of SVM is that the complexity of its computation does not rely on the dimensional parameters of the input space. Moreover, SVR shows better generalization ability, with high performance and accurate prediction.

Fundamentally, SVR is a linear approach with one output, dealing with a space of high dimensional feature established by nonlinear mapping of the N-dimensional input vector into a K- dimensional feature space (K > N) utilizing the function ϕ(x). The learning process is moved to the minimization of the error function, which is defined by the so called ε-insensitive loss function $L_\varepsilon(d, y(x))$:

$$L_\varepsilon(d, y(x)) = \begin{cases} |d - y(x)| - \varepsilon, \; for \; |d - y(x)| \geq \varepsilon \\ 0 \qquad , \qquad for \quad |d - y(x)| < \varepsilon \end{cases} \tag{2}$$

Where, *ε* is the assumed accuracy, *d* is the destination, *x* is the input vector, and *(x)* is the actual output under the effect of x.

The actual output of the SVR is defined by:

$$y(x) = \sum_{j=1}^{K} \omega_j \, \varphi_j(x) + b = w^T \varphi(x) + b \tag{3}$$

Where $w = [\omega_0, \omega_1, \dots, \omega_K]^T$ is the weight vector and $\varphi(x) = [\varphi_0(x), \varphi_1(x), \dots, \varphi_K(x)]^T$ is the basis function vector.

A3. Bayesian Linear Regression Model

Bayesian linear regression is based on a generative method which is different from a discriminant one, which depends on Bayesian inference to build linear regression models [15]. Once the model is specified, the posterior distribution of parameters and forecasts of the model are computed by the method. This statistical analysis enables the method to define the complexity of the model through training, which produces a model with few possibilities to overfit.

In contrast to simple linear regression model, the responses in Bayesian Linear Regression are assumed as samples from the probability distribution, for example the normal (Gaussian) distribution, which is:

$$Y \sim N(\beta^T X, \sigma^2) \tag{4}$$

The product of the parameters *β* and the inputs *X* is the mean of the Gaussian, where the normal deviation is σ. As well as the responses, in Bayesian Models, the parameters are also supposed to be sampled from a distribution. The aim is to define the posterior probability distribution for the parameters of the model with given *X* inputs and *Y* outputs, as in Eq. (5):

$$P(\beta\backslash Y,X)=\frac{P(Y\backslash\beta,X)*P(\beta\backslash X)}{P(Y\backslash X)} \quad (5)$$

The final result obtained from modelling by Bayesian Linear regression is not a single estimate, but rather a distribution range which can be used to produce inferences regarding new observations. This distribution enables determination of uncertainty in the model, which is considered one of the advantages of Bayesian Modelling methods. When the volume of data increases, the uncertainty of the result declines, presenting a better level of certainty in the approximation [17].

A4. Artificial Neural Networks Model

Neural Network or Artificial Neural Network is a mathematical model which is based on the concept of Artificial Intelligence, which simulates the biological neuronal activity of the human brain. This modelling approach is a valuable tool that simulates the functionality of the human brain when dealing with complex relations between the inputs and outputs in any systems [18]. There are many types of ANN architectures, the most common of which is Multi-Layer Perceptron (MLP), which is commonly used for prediction. It comprises three tiers: an input layer, hidden layers, and an output layer. Supposing that the input vector is $\vec{x}$ and the weight vector is $\vec{w}$, and the activation function is a sigmoid function (which is the most commonly used function type), the output is given by:

$$Y=sigmoid(\vec{w}^T.\vec{x}) \quad (6)$$

where the sigmoid(x) is

$$sigmoid(x)=\frac{1}{1+e^{-x}} \quad (7)$$

One of the characteristic advantages of Neural Network technique is its ability to deal with noisy, incomplete, or missing data, requiring no previous assumptions. In addition, it has capabilities to deal with complex relations between input and output variables, and consequently to predict the output of new data input. However, overfitting and overtraining are considered as limitations of Neural Networks. Also, regarding the selection of parameters, in Neural Network there is no formal way to select the suitable parameters for the model, which may influence the accuracy of its prediction.

A5. Adaptive Neuro -Fuzzy Inference Model

The ANFIS model is a combined model of fuzzy systems and ANN [19]. The main parts of the FIS are fundamental rules, which contain the choices of fuzzy logic rules “If-Then”, a set of membership functions, and the fuzzy logic inference procedures from the fundamental rules to obtain the output. In order to map the inputs with the outputs, two common fuzzy inference systems (FIS) can be employed in different applications: Mamdani and Sugeno inference systems.

The fuzzy rules in the two inference models give different results, therefore their actions of defuzzification and combination are also different. However, the Sugeno system is believed to be computationally more efficient than the Mamdani; in the former, the resultant parameter is a linear equation or constant coefficient. Supposing that we have a system including two inputs, x and y, and the output is f, and the based rule has two fuzzy if-then rules, then the description of rules for the linear equation Sugeno FIS can be presented as the rule 1 (R1) and rule 2 (R2):

R1: if $x\ is\ A_1$ and $y\ is\ B_1$ then $f_1=p_1x+q_1x+r_1$ (8)

R2: if $x\ is\ A_2$ and $y\ is\ B_2$ then $f_2=p_2y+q_2y+r_2$ (9)

where A_i and B_i are the membership functions of each input x and y; and p_i, q_i and r_i are the linear parameters in the resulting part of the Sugeno fuzzy inference system.

ANFIS model can be consider as a successful model due to the strength of its results. Moreover, as with other machine learning techniques and as a neural network, ANFIS has a high ability to generalize. On the other hand, there are some limitations of ANFIS model regarding the type, number, and position of membership functions [20].

B. Combining Models

Combining a set of models or classifiers is an attractive and common way of trading-off the strengths and limitations of various machine learning and statistical techniques [21]. After obtaining a set of trained and tested models, instead of choosing the best-performing model, it is valuable to apply ensemble methods to combine the predictions of all these models, in order to get the most accurate and reliable model [22][23].

There are several available methods of combination, the most simple and popular of which are Majority vote (MAJ), Maximum (MAX), Minimum (MIN), Average (AVR), and Product (PRO). In this study the following methods have been applied: simple average, weighted average, and majority vote [24].

B1. Simple Averaging

Simple averaging (AVR) rule is among the most basic and popular combination methods for numerical values. This approach is simple to apply as there is no need for previous training [22][25]. This rule combines the outputs of multiple models by taking the average directly. If we have a set of B single learners $\{h_1, h_2, \ldots \ldots h_B\}$ and the output of h_i, for instance x is $h_i(x) \in R$, the combined outputs can be defined as:

$$C(x) = \frac{1}{B}\sum_{i=1}^{B} h_i(x) \tag{10}$$

B2. Weighted Averaging

Weighted averaging (WAVR) is an extension of the simple averaging rule. In this method, the outputs of all models are combined by taking the average with different weights indicating different levels of importance [21]. In general, the combined output can be defined as:

$$C(x) = \sum_{i=1}^{B} w_i\, h_i(x) \tag{11}$$

where w_i is the weight for h_i, and the weights w_i are normally supposed to be constrained by $w_i \geq 0$ and:

$$\sum_{i=1}^{B} w_i = 1 \tag{12}$$

B3. Majority Voting

Majority voting (MAJ) is considered as one of the most common ways of voting in statistical analysis. In this method, the predictions of multiple models are combined. The predictions of every single model is represented as a single vote, and the final output is the one that obtains the majority votes of the models [21].

If there are three different models, $h_1(x), h_2(x), h_3(x)$, for a specific classification or regression problem, these models can be combined as:

$$C(x) = mode\{h_1(x), h_2(x), h_3(x)\} \tag{13}$$

Generally, a majority vote combiner involving votes from multiple learners (models) $h_1, h_2, \ldots \ldots h_B$, can be defined as:

$$C(x) = arg \max_i \sum_{j=1}^{B} (h_j(x) = i) \tag{14}$$

IV. EXPERIMENTAL METHODOLOGY

Data for the prevalence of diabetes, smoking, obesity, and inactivity in SA were obtained from Saudi Health Interview Survey [26], which is provided by the Saudi Ministry of Health, along with other published national surveys [27][28][29][30][31]. These community-based national studies include adults aged 15 years and over. In addition, the diagnostic criteria used as a diabetes detection method was either WHO or American Diabetes Association (ADA) criteria. In this study, obesity as a risk factor was defined according to the definition of body mass index (BMI ≥ 30 kg/m^2), and for smoking only data for current smokers was taken.

The data is arranged according to age and gender (demographic factors). Data were divided into six ten-year age bands (25-34, 35-44, …. …, +75 years old) for men and women.

For the missing data between the years, we used ANFIS model to calculate them and we did the same calculations for the future data from 2013 up to 2025.

It is valuable to point out that building a good predictive model requires a good choice of input variables. The choice of these inputs should be such that the functions of the model perform accurately between the inputs and outputs.

The morbidity data of diabetes considered as a dependent variable, whereas risk factors data (age, gender, smoking, obesity, inactivity) are used as independent variables. The purposed models were deployed in MATLAB version 2018a.

After training the data through the models, we combined the outputs of these models using three different methods of combination (AVR, WAVR, MAJ). Tables I and II show the results of combining methods for men and women, respectively.

TABLE I. MODELS AND COMBINED RESULTS (MALE TEST DATA)

	Models Results					**Combiners**		
Year	**MLR**	**ANFIS**	**ANN**	**SVR**	**BLM**	**AVR**	**WAVR**	**MAJ**
2014	14.1	14.2	14.5	13.8	14.4	14.2	14.1	14
2015	14.2	14.4	15.1	13.9	14.7	14.5	14.3	14
2016	14.4	14.8	15.8	14.0	15.1	14.8	14.5	14
2017	14.6	15.2	16.6	14.1	15.5	15.2	14.8	14
2018	14.7	15.5	17.0	14.1	15.9	15.5	14.9	14
2019	14.9	16.0	17.7	14.2	16.3	15.8	15.2	15
2020	15.1	16.6	18.2	14.2	16.7	16.1	15.4	15
2021	15.2	17.1	18.6	14.3	17.1	16.5	15.6	15
2022	15.4	17.7	18.8	14.3	17.5	16.7	15.8	15
2023	15.5	18.3	18.9	14.4	17.9	17.0	15.9	14
2024	15.7	19.0	19.1	14.4	18.3	17.3	16.1	19
2025	15.9	19.6	19.2	14.5	18.8	17.6	16.2	14

TABLE II. MODELS AND COMBINED RESULTS (FEMALE TEST DATA)

	Models Results					**Combiners**		
Year	**MLR**	**ANFIS**	**ANN**	**SVR**	**BLM**	**AVR**	**WAVR**	**MAJ**
2014	12.9	12.2	11.5	10.7	12.6	11.9	12.8	12
2015	13.4	12.8	11.7	10.8	13.1	12.3	13.3	13
2016	14.6	13.1	12.3	10.9	14.2	13.0	14.3	15
2017	15.1	13.8	13.1	11.0	14.6	13.5	14.9	15
2018	15.7	14.4	14.1	11.1	15.1	14.1	15.4	14
2019	16.8	14.8	15.5	11.2	16.1	14.9	16.2	15
2020	16.8	15.7	15.9	11.4	16.0	15.1	16.5	16
2021	17.3	16.4	16.4	11.5	16.5	15.6	17.1	16
2022	17.9	17.0	16.8	11.6	16.9	16.0	17.6	17
2023	19.0	17.4	17.0	11.7	18.0	16.6	18.6	17
2024	19.5	18.1	17.1	11.8	18.4	17.0	19.2	20
2025	20.1	18.7	17.2	12.0	18.9	17.3	19.7	20

V. RESULTS AND DISCUSSION

As illustrated in Fig.1, the prevalence rate of diabetes was estimated to rise from 8.4% in 1999 to 17.5% by 2025. Comparing the prevalence of diabetes according to different demographic factors, such as gender and age, showed that the prevalence rate was higher among men than women; 9.7% in men compared to 7% in women in 1999, projected to reach 17.6% for men and 17.4% for women by 2025. Furthermore, it varied according to age group; for both men and women, the prevalence was lower in younger age groups, and it increased with age (Fig. 2, Fig. 3). Also, the cohort aged 55-64 years old had the significant prevalence rate, while it was slightly lower among those aged 65 years and over.

The results in Fig. 4 show that the prevalence of the behavioral risk factors smoking and obesity also increased from 1999 to 2025: smoking increased from 11% to 16.05%, while obesity sharply increased from 16.7% to 51.7%. The prevalence rate of inactivity is expected to have dropped significantly to 61.1% by 2025, compared to 96% in 1999, however this percentage remains dangerously high. According to [32], the author indicated that “Saudis are not active enough to meet the recommended guidelines for moderate to vigorous PA”, which has been attributed to the rapid increase of urbanization, the nature of weather, and cultural characteristics.

In addition, the prevalence of risk factors including obesity, smoking, and physical inactivity varied according to gender. The prevalence of smoking was higher among men than women: 21.1% against 0.9% in 1999, and 28.4% against 3.7% by 2025. Women had a higher prevalence of obesity than men: 20.3% against 13.1% in 1999, and 58.4% against 45% by 2025 (Fig. 4). Moreover, the prevalence rate of physical inactivity was higher among women than men: 98.1% against 93.9% in 1999, and 71.7% against 50.5% by 2025.

Generally, demographic as well as behavioural risk factors significantly contributed to the increased level of diabetes, with a significant level of 0.05, however smoking, obesity and physical inactivity were the most significant factors (Table III).

TABLE III. THE RELATIONSHIP BETWEEN DIABETES PREVALENCE AND RELATED RISK FACTORS.

Variables	P-value
Gender	0.02
Age	0.01
Smoking	0.000
Obesity	0.001
Inactivity	0.001

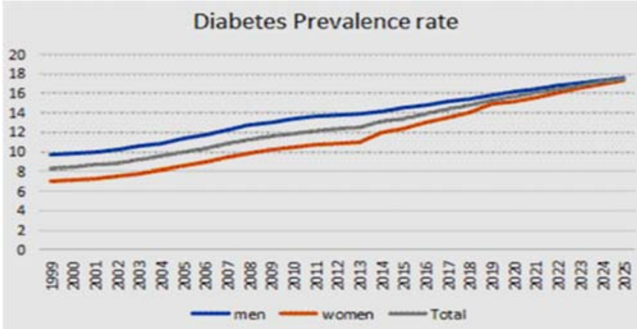

Fig. 1. Diabetes prevalence estimations for Saudis aged 25-75+, 1999-2025.

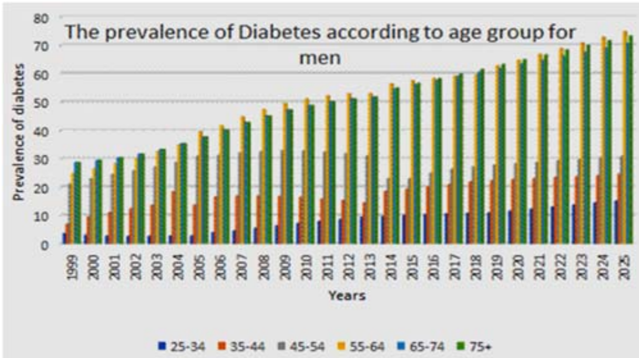

Fig. 2. Diabetes prevalence estimations for men according to age groups.

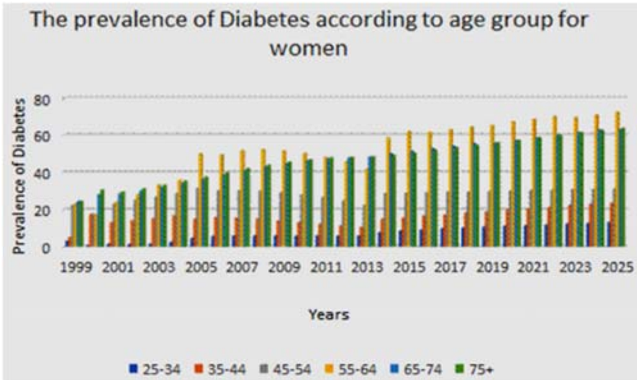

Fig. 3. Diabetes prevalence estimations for women according to age groups.

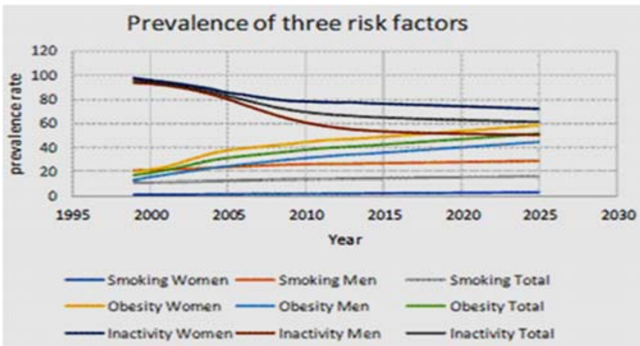

Fig. 4. Prevalence rates of smoking, obesity, and inactivity for Saudis aged 25-75^{+}, 1999-2025.

In order to validate the combined models, results are compared with other estimated findings from different studies of diabetes prevalence in the SA for years 2015 to 2019, as shown in Table IV [8][33][34][35][36].

TABLE IV. A COMPARISON OF THE COMBINED MODEL AGAINST OTHER STUDIES OF DIABETES PREVALENCE IN SAUDI ARABA, 2015-2019.

Study	Diabetes prevalence estimates	Combined model
Saudi Arabia Health Profile WHO (2015)	Men: 14.8 Women: 11.7 Total: 13.3	Men: 14.5 Women: 12.4 Total: 13.5
Diabetes Country Profile: Saudi Arabia WHO (2016)	Men: 14.7 Women: 13.8 Total: 14.3	Men: 14.8 Women: 13.02 Total: 13.9
Family Health Survey Saudi General Authority for Statistics (2017)	Men: (10.4) *, Men: (14.9) ** Women:(9.8) *, Women:(14.5)** Total: (10.1) * , Total:(14.7)**	Men: 15.2 Women: 13.5 Total: 14.4
Household Health Survey Saudi General Authority for Statistics (2018)	Men: (10.3) *, Men: (16.8) ** Women:(9.9) *, Women:(14.2) ** Total: (10.1) *, Total:(15.5) **	Men: 15.5 Women: 14.1 Total: 14.8
Diabetes Atlas (9th edition), IDF (2019)	Total: 15.8 (10.3-17.7) (95% confidence interval)	Men: 15.8 Women: 14.9 Total: 15.4

* Prevalence rate for Saudi population (15 years and over)
** Prevalence rate for Saudi population (age-adjusted 25 years and over)

Overall, according to our validation results, simple averaging resulted in slightly better performance than weighted averaging and majority voting. Thus, after comparing the performance of our combined model with the individual models (independently), we find that ANFIS model has the most similar results with the combined model, indicating greater accuracy.

VI. CONCLUSION AND FUTURE WORK

This paper has developed and integrated different mathematical models in order to study the prevalence rate of diabetes and its related factors, aiming to support health policy makers and planners for programming and resource allocation. These models were applied on diabetes data obtained from Saudi Arabia, where the level of diabetes prevalence is predicted to increased affected by the increased population and ageing, along with rising levels of different risk factors. We found that smoking, obesity and physical inactivity were the most significant contributing factors in diabetes. These results highlight the urgency need for seek more strategies and preventive programmes to overcome this disease and reduce its complications, or at least delay them.

The proposed modelling approaches have been combined and validated against other results obtained from different studies. The most accurate model was ANFIS model, which also had the smallest RSME, making it the most appropriate and accurate of the studied models.

REFERENCES

[1] N. A. Rosenthal, "Infections, Chronic Disease, and the Epidemiological Transition: A New Perspective," *Clin. Infect. Dis.*, vol. 61, no. 3, pp. 489–490, Aug. 2015.

[2] J. Frenk, J. L. Bobadilla, J. Sepuúlveda, and M. L. Cervantes, "Health transition in middle-income countries: New challenges for health care," *Health Policy Plan.*, vol. 4, no. 1, pp. 29–39, 1989.

[3] A. Esteghamati *et al.*, "Third national Surveillance of Risk Factors of Non-Communicable Diseases (SuRFNCD-2007) in Iran: methods and results on prevalence of diabetes, hypertension, obesity, central obesity, and dyslipidemia," *BMC Public Health*, vol. 9, p. 167, 2009.

[4] C. S. Cockram, "The epidemiology of diabetes mellitus in the Asia-Pacific region," *Hong Kong Med. J.*, vol. 6, no. 1, pp. 43–52, 2000.

[5] World Health Organization, "WHO South-East Asia," *WHO South-East Asia J. Public Heal.*, vol. 5, no. 2, pp. 77–173, 2016.

[6] A. T. Kharroubi, "Diabetes mellitus: The epidemic of the century," *World J. Diabetes*, vol. 6, no. 6, p. 850, 2015.

[7] S. Wild, G. Roglic, A. Green, R. Sicree, and H. King, "Global Prevalence of Diabetes," *Diabetes Care*, vol. 27, no. 5. pp. 1047–1053, 2004.

[8] I. D. F. D. Atlas, "International Diabetes Federation. IDF Diabetes Atlas, 9th edn. Brussels, Belgium," 2019.

[9] A. J. Palmer *et al.*, "Computer modeling of diabetes and its complications: a report on the Fourth Mount Hood Challenge Meeting.(ADA WORKGROUP REPORT)(Conference notes)," *Diabetes Care*, vol. 30, p. 1638, 2007.

[10] S. F. U. C. S. M. Group, *Modelling in Healthcare*. American Mathematical Society, 2010.

[11] D. R. Whiting, L. Guariguata, C. Weil, and J. Shaw, "IDF diabetes atlas: global estimates of the prevalence of diabetes for 2011 and 2030," *Diabetes Res. Clin. Pract.*, vol. 94, no. 3, pp. 311–321, 2011.

[12] L. Guariguata, D. R. Whiting, I. Hambleton, J. Beagley, U. Linnenkamp, and J. E. Shaw, "Global estimates of diabetes prevalence for 2013 and projections for 2035," *Diabetes Res. Clin. Pract.*, vol. 103, no. 2, pp. 137–149, 2014.

[13] "Erratum: Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants (The Lancet (2016) 387(10027) (1513–1530)(S0140673616006188)(10.1016/S0140-6736(16)00618-8))," *The Lancet*, vol. 389, no. 10068. p. e2, 2017.

[14] T. P. Ryan, *Modern Regression Methods*. Wiley, 2008.

[15] M. Awad and R. Khanna, *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*. Apress, 2015.

[16] R. Sałat and K. Sałat, "The application of support vector regression for prediction of the antiallodynic effect of drug combinations in the mouse model of streptozocin-induced diabetic neuropathy," *Comput. Methods Programs Biomed.*, vol. 111, no. 2, pp. 330–337, 2013.

[17] W. Koehrsen, "Bayesian Linear Regression in Python: Using Machine Learning to Predict Student Grades Part 2," *towards data science*, 2018. [Online]. Available: https://towardsdatascience.com/bayesian-linear-regression-in-python-using-machine-learning-to-predict-student-grades-part-2-b72059a8ac7e%0D.

[18] W. Suparta and K. M. Alhasa, *Modeling of Tropospheric Delays Using ANFIS*. Springer International Publishing, 2015.

[19] L. Rutkowski, *Flexible Neuro-Fuzzy Systems: Structures, Learning and Performance Evaluation*. Springer US, 2006.

[20] M. N. M. Salleh, N. Talpur, and K. Hussain, "Adaptive Neuro-Fuzzy Inference System: Overview, Strengths, Limitations, and Solutions BT - Data Mining and Big Data," 2017, pp. 527–535.

[21] Z. H. Zhou, *Ensemble Methods: Foundations and Algorithms*. CRC Press, 2012.

[22] S. B. Kotsiantis and P. Pintelas, "Selective averaging of regression models," *Ann. Math. Comput. Teleinformatics*, vol. 1, no. 3, pp. 65–74, 2005.

[23] M. Re and G. Valentini, "1 Ensemble methods: a review 3," 2012.

[24] M. Ponti, *Combining Classifiers: From the Creation of Ensembles to*

the Decision Fusion. 2011.
[25] C. A. Shipp and L. I. Kuncheva, "Relationships between combination methods and measures of diversity in combining classifiers," *Inf. fusion*, vol. 3, no. 2, pp. 135–148, 2002.
[26] "Saudi Health Interview Survey Results," 2013.
[27] A. S. Warsy and M. A. El Hazmi, "Diabetes mellitus, hypertension and obesity-common multifactorial disorders in Saudis," 1999.
[28] M. M. Al-Nozha *et al.*, "Diabetes mellitus in Saudi Arabia.," *Saudi Med. J.*, vol. 25, no. 11, pp. 1603–1610, 2004.
[29] "WHO STEPwise Approach to NCD Surveillance, Country-Specific Standard Report, Saudi Arabia;," 2005.
[30] M. M. Al-Nozha *et al.*, "Obesity in Saudi Arabia.," *Saudi Med. J.*, vol. 26, no. 5, pp. 824–829, 2005.
[31] J. S. Jarallah, K. A. Al-Rubeaan, A. R. A. Al-Nuaim, A. A. Al-Ruhaily, and K. A. Kalantan, "Prevalence and determinants of smoking in three regions of Saudi Arabia," *Tob. Control*, vol. 8, no. 1, pp. 53–56, 1999.
[32] H. M. Al-Hazzaa, "Physical inactivity in Saudi Arabia revisited: a systematic review of inactivity prevalence and perceived barriers to active living," *Int. J. Health Sci. (Qassim).*, vol. 12, no. 6, p. 50, 2018.
[33] R. O. for the E. M. World Health Organization, *Saudi Arabia health profile 2015*. Cairo PP - Cairo: World Health Organization. Regional Office for the Eastern Mediterranean.
[34] "World Health Organization – Diabetes country profiles," 2016. [Online]. Available: www.who.int/diabetes/country-profiles/sau_en.pdf.
[35] (General Authority For Statistics) Saudi Arabia, "Family health survey," 2017.
[36] (General Authority For Statistics) Saudi Arabia, "Household Health survey," 2018.

.