SINDH UNIVERSITY RESEARCH JOURNAL (SCIENCE SERIES)
ISSN (Print) 1813-1743, ISSN (online) 2791-0547
Vol. 57:01 (2025).
Doi: https://doi.org/10.26692/surjss.v57i1.7561

AI-Powered Genetic Algorithm-Based System for Predicting Diabetics Risk in Future Generations

Syeda Azwa Asif [1]*, Syed Asif Ali [2], Asma Khan[1],

[1]Department of Software Engineering, NED University of Engineering and Technology, Karachi, Pakistan
[3]University Road, , NED University of Engineering and Technology, Karachi, Pakistan
[2]Sindh Madressatul Islam University Karachi, Pakistan/Aiwan-e-Tijarat Rd, Seari Quarters, Karachi, 74000.

Cite this:
Syed Asif Ali, Syeda, Azwa, Asma Khan. AI-Powered Genetics Algorithm-Based System for Predicting Diabetics Risk in Future Generations Uni. Res.J. (SS) 57. 01. 2025

Corresponding author
syedaazwa@gmail.com
orcid: 0009-0008-7103-3279

ABSTRACT

Diabetes is a global health concern leading to complications such as high blood pressure, kidney failure, and vision impairment. Effective management of diabetes involves maintaining a healthy lifestyle with proper diet and exercise. This research focuses on predicting the transmission of diabetes to future generations using classical Mendelian genetics combined with a probabilistic model enhanced by Artificial Intelligence (AI) and machine learning (ML). Real-time data was collected from endocrinologist associated with various families maintaining the ethical privacy of their datasets. The study employs Mendelian laws of genetics to analyze the inheritance patterns of diabetes focusing on dominant and recessive alleles. Integrating Mendelian genetics with probabilistic rules (Naïve Bayes), the methodology offers a robust framework for predicting diabetes transmission. The findings highlight significant generational differences, revealing that diabetic parents have a strong likelihood in a sample of 400 individuals of passing diabetes to their offspring. This intelligent application allows individuals without technical or biomedical knowledge to easily predict the risk of diabetes in their next generation. This application validated through Mendelian laws and probabilistic models using AI and ML, holds promise for future expansion to address related diseases like kidney failure, eye problems, high blood pressure, and heart disease. By making this tool accessible to a lay audience, it enhances understanding of diabetes inheritance and offers a powerful resource for genetic prediction.

Keywords: Diabetes, Mendelian genetics, type 2 diabetes, artificial intelligence, machine learning

INTRODUCTION

Diabetes is an ongoing health condition that arises when the pancreas is unable to produce enough insulin or if the body fails to utilize the insulin that is produced. The purpose of insulin as a hormone is to maintain the level of glucose in the blood. It is a common disease affecting people of all ages.

Medically, there are many types of diabetes, but two are very common: type one diabetic (TOD or type-I) and type two diabetic (T2D or type-II). Type 2 diabetes can be inherited. The risk is higher in the offspring or upcoming generation if the mother, rather than the father, has T2D. Predicting the risk in the next generation or offspring and timely intervention are crucial. Most diabetic forms are lifelong but manageable (World Health Organization,2023). The proposed application aims to predict diabetes in future generations through a blend of classic Mendelian genetics. Utilizing data collected from run-time inputs, a diverse dataset captures real-world family scenarios. Applying Mendelian genetics to input data sets, it explores how traits like diabetes are inherited through generations considering dominant and recessive alleles. Simultaneously, the probability law (Naïve Bayes) incorporates data and historical data, including parental diabetic status and previous history for a probabilistic prediction (IDF,2023), (Prasad, R. B. et al.,2015), (Gautier, T. et al.,2021).

Table 1: Statistic of Diabetes in Pakistan (Mujumdar, A. et al.,2019)

S. No.	Years	Diabetes statistic
1	2000	11.8%
2	2016	11.77%
3	2018	16.9%
4	2019	17.1%
5	2022	26.7%
6	2045	33.6%

BACKGROUND RESEARCH

The frequency of diabetes is increasing day by day. According to the International Diabetes Federation (IDF), Pakistan is in the 3rd position being most affected by diabetes. The rising prevalence of diabetes isn't limited to adults; it's also affecting children. Thus, in pursuit of the goal of detecting the probability of diabetes in the next generation, the decision is made to propose an application predicting the probability of diabetics in the upcoming or next generation (Priya, K. L. et al., 2020), (Pakistan diabetes report, 2000).

This research covers diabetic aspect along with AI, which creates bridging between medical and advanced technology field. The research in these areas is being conducted which are discussed as followed.

Birjais, R. et al. (2019), in research titled, "Predicting Diabetes Mellitus with Machine Learning Techniques", discusses fasting glucose as a key indicator used for diabetes prediction and other factors which are needed for accuracy. Random forests outperformed other classifiers. The best accuracies were 0.8084 (Luzhou) and 0.7721 (Pima Indians). Future work will focus on predicting diabetes types and improving accuracy by analyzing indicator proportions (Birjais, R. et al.,2019).

In 2018, Sisodia, D. et al., suggests in research paper, "Prediction of Diabetes using Classification Algorithms" a system for early-stage diabetes detection using three machine learning algorithms evaluated on the Pima Indians Diabetes Database. The Naive Bayes algorithm achieved 76.30% accuracy. The system can potentially be extended to diagnose other diseases and improved for automated diabetes analysis with additional algorithms (Sisodia, D. et al.,2018).

Kavakiotis, I. et al. (2017), in research titled, "Machine Learning and Data Mining Methods in Diabetes Research" reviewed machine learning and data mining applications in diabetes mellitus research, emphasizing its global health impact. Significant advancements have been made in biomarker identification and prediction-diagnosis. Future exploration is expected to leverage biotechnology and extensive electronic health records for deeper insights into DM's diagnosis, pathophysiology, and treatment sing advanced analytics (Kavakiotis, I. et al.,2017).

Tapak, L. et al. (2013), in research titled, "Real-Data Comparison of Data Mining Methods in Prediction of Diabetes in Iran" concludes that the support vector machine (SVM) model outperformed other classifiers in predicting diabetes, showing superior sensitivity, specificity, and overall accuracy. This highlights SVM as a promising tool not only for diabetes prediction but also for potential application in predicting other diseases (Tapak, L. et al.,2013).

Yu, W. et al. (2010), in research titled, "Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes" proposed a promising classification method for identifying individuals within the population with common diseases like diabetes and pre-diabetes in support with vector machine modeling. Using common variables, this strategy should be investigated further in other complex disorders (Yu, W. et al.,2010).

OBJECTIVE

As mentioned in the above session, the rate of diabetes is increasing day by day so there is a need for a system to predict diabetes in upcoming generations to mitigate diabetes and related problems. The objective of this research is to help prevent diabetes and its associated diseases including, eyes related issues, kidney failure, blood pressure and heart related disease by developing a software application to predict the transmission of diabetes in upcoming generations utilizing classical Mendelian genetics along with a probabilistic model enhanced by AI, and ensuring the application has a user-friendly and attractive graphical user interface (GUI) so that non-technical individuals can easily operate the application.

INVESTIGATING DIABETIC GENETIC MUTATIONS

Genetic mutations play a crucial role in determining variations in diabetes. This involves a thorough exploration of the genetic mutations associated with parental history. Investigating these mutations, helps in unraveling the underlying genetic complexities that increases the probability of being diabetic in human populations.

GENETICS AND HUMAN INHERITANCE

Genetics is the study of living things related to the gene-based history of inheritance transfer from parents to their upcoming generations. It is passed through the human offspring and helps in determining characteristics like resemblance in the face, eyes, or disease that may transfer genetically, like diabetes, thalassemia, etc. Human genetics follows the patterns of inheritance (one copy comes from the mother side as inherited and the second from the father side) and molecular techniques for identifying the interaction between genes and environmental factors, followed by the Mendelian laws of inheritance (genetics) (Xu, S., 2022) , (Germain, D. P. et al.,2018).

THE ROLE OF MENDELIAN GENETICS

Mendelian genetics, introduced by Gregor Mendel, provides a foundation for understanding inheritance patterns. This inheritance pattern helps in verifying the applicability of Mendelian principles for diabetics in an inheritance. By confirming or refining these principles, we can enhance our understanding of how genetic information is passed from one generation to the next, specifically in the context of diabetic probability (Zschocke, J. et al.,2023).

Mendel proposed three laws of inheritance. The laws of dominance, segregation, and independent assortment (Bateson, W. et al.,2013), (Yule, G. U.1902), (Admin. 2023).

- The law of dominance states, "few alleles are dominant and others are recessive."
- The law of independence shows, "how the different genes of different alleles assort into gametes"
- The law of segregation shows that "the same genes of alleles assort into different gametes."

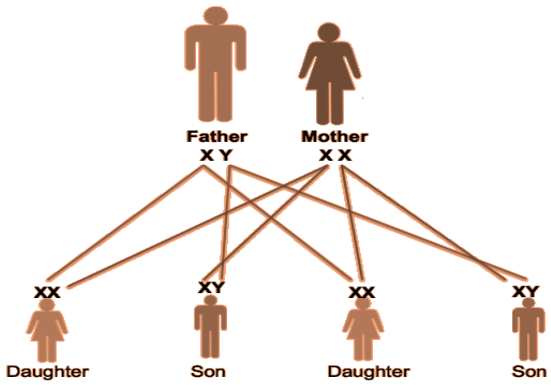

Figure 1: Mendelian Genetics

NAÏVE BAYES ALGORITHM

Integrating a probabilistic approach into genetic studies represents a novel concept. The Naïve Bayes algorithm is a probabilistic machine learning algorithm based on the Bayes Theorem. This theorem is grounded on the probability of a hypothesis, given the data and some prior knowledge (Priya, K. L., et al.,2020).

$$P(A \mid B) = \frac{P(B|A) \times P(A)}{P(B)} \quad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

where;

$P(A \mid B) \rightarrow$ Posterior probability of class A given predictor B.

$P(B \mid A) \rightarrow$ Likelihood, the probability of observing predictor B given that class A is true.

$P(A) \rightarrow$ Prior probability of class A.

$P(B) \rightarrow$ Marginal probability of predictor B.

The Naive Bayes algorithm in machine learning is very beneficial in classification problems used for categorizing uninitiated data from predefined classification. The algorithm learns from patterns in existing data, offering a computational perspective on predicting complex phenotypic traits. Leveraging this technology provides a more nuanced and probabilistic understanding of diabetes inheritance (Berrar, D.,2019). In machine learning during the training of Naïve Bayes algorithm, probabilities for all possible combinations occurrence of diabetes in upcoming generation based on their past genetic history of diabetes (parents) and prediction for future generation values are calculated and stored category wise. Testing this algorithm repossesses, all probability based on future generation diabetes prediction helps in achieving final output. The precalculated probability from history helps in getting efficient results.

PUNNETT SQUARE

Punnett introduced the concept of "Punnett Square" to represent the number and variety of genetic combinations. The Punnett square is a simple graphical or tabular form that shows all possible genotypes (crosses between two individuals) from a particular cross as shown in figure 2. The Punnett square is a pictorial tool or technique that is used to predict the genotype of a cross or determine the probability of an upcoming generation having a particular type of genotype (Sebastiani, P. et al.,2012), (Rossi, A., et al., 2022).

The simplest form of Punnett square consists of a square divided into four equal quadrants, which label the rows with one parent genotype and the column with the genotype of another parent. There are two major types of Punnett squares. The first type is called monohybrid; it's a single genetic trait inherited by crossing two homozygous (dominant and recessive) parents. Monohybrid is a 2×2 square composed of four boxes. The second type of Punnett square is a dihybrid (codominant) trait, which is a 4×4 square with 16 boxes and deals with more than one genetic trait (Davis, L. C.,1993), (Müller-Wille, S. et al.,2020), (Edwards, A. W. F.,2012).

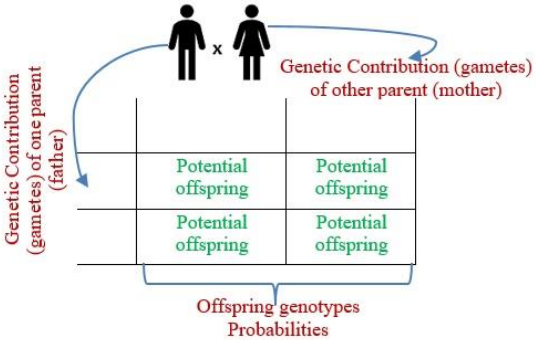

Figure 2: Punnett Square

Using a monohybrid cross, in which the recessive allele is represented by d and the dominant allele by D:

- Homozygous dominant → DD
- The heterozygous → Dd
- The homozygous recessive → dd

Table 2: Punnett Square		
	D	**d**
D	DD	Dd
d	Dd	dd

The following are the possible 16 combinations that can be made from Punnett square. The probability of diabetes transmission in next generations can be calculated by Bayes Algorithm.

Table 3: Possible 16 Combination by Punnett Square				
S. No	Father	Mother	Father History	Mother History
1	0	0	0	0
2	0	0	1	0
3	0	0	0	1
4	0	0	1	1
5	1	1	0	0
6	1	1	1	0
7	1	1	0	1
8	1	1	1	1
9	0	1	0	0
10	0	1	1	0
11	0	1	0	1
12	0	1	1	1
13	1	0	0	0
14	1	0	1	0
15	1	0	0	1
16	1	0	1	1

GENOTYPE

A genotype is the unique sequence of human DNA used to represent the two forms; a human has inherited from their parents for a specific gene. The phenotype is the most noticeable expression of this genotype (Zschocke, J. et al.,2023),(Genetics, autosomal dominant, PubMed,2024).

Basically, the risk of diabetes in the upcoming generation is usually influenced by a single gene with two alleles, which oversees the hereditary weakness associated with the condition.

- **D** = Dominant allele (diabetes risk)
- **d** = Recessive allele (no diabetes risk)

Table 4: Genotypes with probability				
S. No	genotypes	Risk	Description	Probability
1	dd	High Risk	Both parents contribute the dominant D allele	25%
2	Dd	Moderate Risk	one parent contributes D, the other d	50%
3	DD	Low Risk	both parents contribute the recessive d allele	25%

MATERIALS AND METHODS

Approach for Data Sets

In order to identify and predict diabetes in next generation, real-time data is collected from endocrinologists with the consent of the hospital, associated with various families, maintaining the ethical privacy of their datasets. Major demographic variables included age, gender, diabetic status, diagnostic age, diabetic history, etc. At the initial level, data set of 400 diabetic affected families was collected for designing the application and testing. For research purpose, representative subset of the data set is used (Table 5).

Table 5: Major Parameters of Diabetics with random data sets					
Individual gender	Age	Marital status	Diabetic	Past History	Number of children
Male	41	Yes	Yes	Yes	3
Female	32	Yes	Yes	Yes	2
Female	35	Yes	Yes	Yes	4
Female	50	Yes	Yes	Yes	3
Male	48	Yes	Yes	Yes	2
Male	34	Yes	Yes	Yes	2
Male	51	Yes	Yes	Yes	1
Female	54	Yes	Yes	No	2
Female	23	Yes	Yes	No	3
Male	35	Yes	Yes	No	5
Male	45	Yes	Yes	No	3
Male	47	Yes	Yes	Yes	4
Male	55	Yes	Yes	Yes	6
Male	29	Yes	Yes	No	1
Male	35	Yes	Yes	Yes	1
Female	32	Yes	Yes	Yes	2
Female	32	Yes`	Yes	Yes	3
Female	33	Yes	Yes	No	4
Male	24	Yes	Yes	Yes	1
Male	38	Yes	Yes	No	0

MODELS AND ALGORITHMS FOR APPLICATION

Step 1 – Importing libraries and dataset
Step 2 – Data Classification
Step 3 – Training Naïve Bayes model on the training set
Step 4 –Testing and prediction
Step 6 – Display Probability visually (graph)

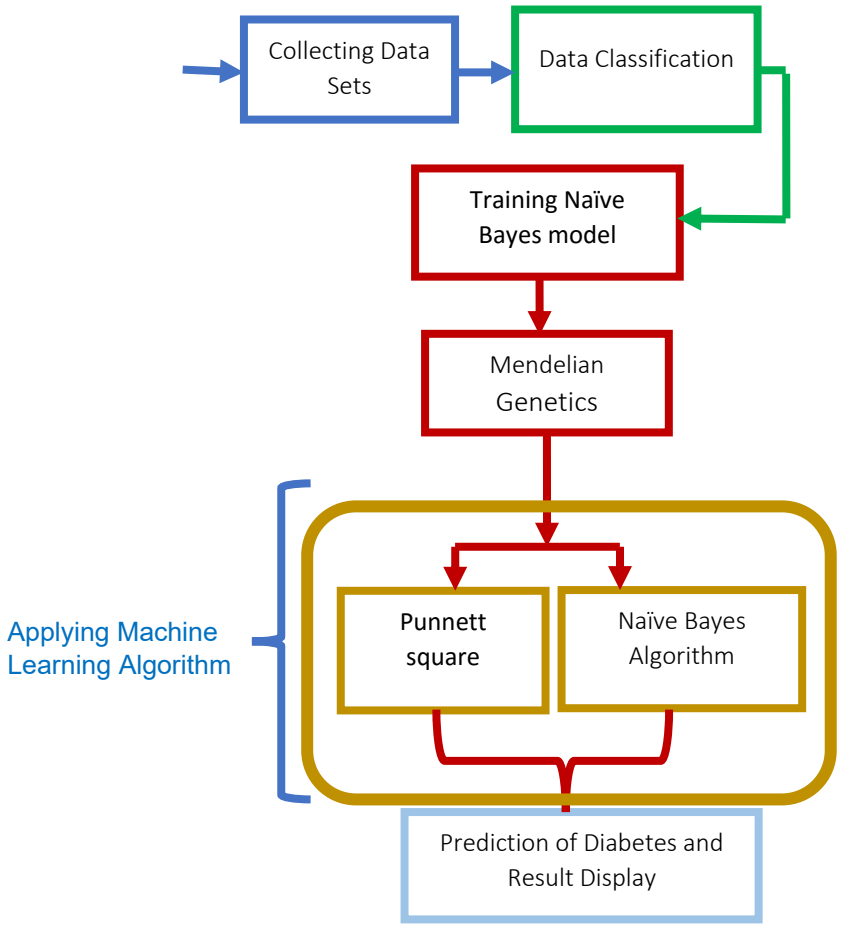

Figure 3: Working Model

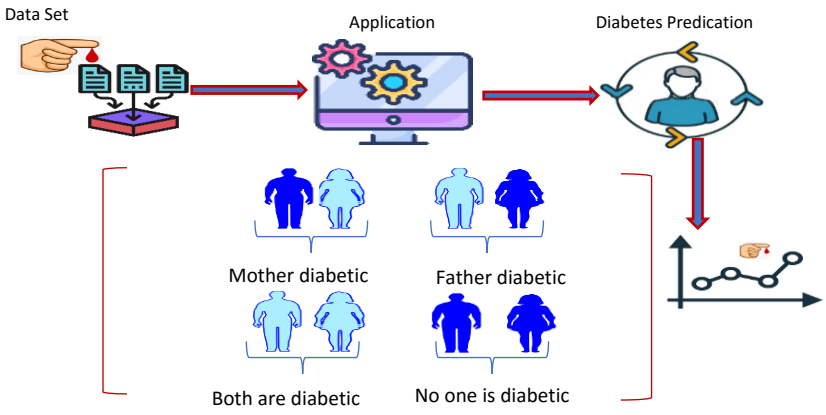

Figure 4: Conceptual Model of proposed

DATA ANALYSIS AND INTERPRETATION

The analysis focuses on predicting the genetic passing of diabetes in future generations based on Mendelian genetics and AI-powered genetic algorithms. The scenario below evaluates the 4 offsprings of a carrier father (whose mother is diabetic) and a non-diabetic mother (whose family has no diabetic history).

Table 6: Case Summary

Grand father (paternal side)	Grand mother (paternal side)	Father	Mother	children
non diabetic	diabetic	carrier	non diabetic	4

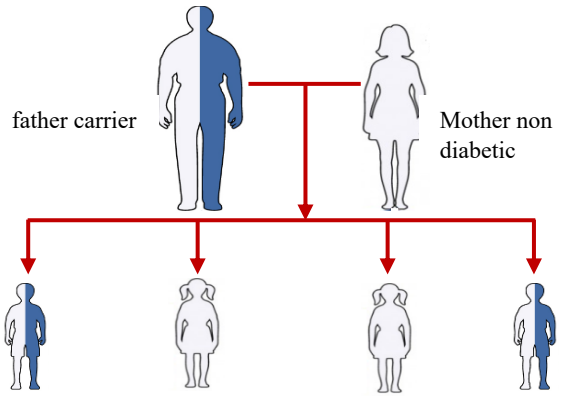

Figure 5: father is carrier and mother is non diabetic having four children. two will be carrier and two are non-diabetic (gender can be varied. Similarly, it can be different as the number of children increases.)

TABLE 7: Case Punnett table

FATHER \ MOTHER	D	D
D	DD (NON-DIABETIC)	DD (NON - DIABETIC)
D	DD (CARRIER)	DD (CARRIER)

GENETIC BACKGROUND

- The father is a carrier (Dd), meaning he carries one dominant allele (D) and one recessive allele (d) for diabetes.
- The mother is non-diabetic and genetically homozygous dominant (DD), meaning she does not carry the diabetic allele.
- The father's father (grandfather) is DD (non-diabetic), and the father's mother is diabetic (dd).

Using Mendelian genetics to predict the offspring's genotypes, and a Punnett square will illustrate the possible combinations of alleles:

Punnett Square Outcomes:

- Probability of having a non-diabetic child (DD) is 50% (since all offspring have at least one D allele).
- Probability of having an offspring who is a carrier (Dd) is 50%.
- 0% chance of having a diabetic (dd) offspring, since both parents do not carry two d alleles.
- While no offspring in this generation will be diabetic (no dd genotype), carriers (Dd) can still pass the recessive allele (d) to their offspring. If a Dd carrier child from this generation has offspring with another carrier (Dd), there is a 25% chance of producing a diabetic offspring (dd) in the next generation.

Thus, the prediction is that the children will not exhibit diabetes, though they can inherit the carrier status (Dd) and potentially pass it to future generations.

Bayes' Theorem for Probability Calculation:

To calculate the probability of an offspring being diabetic or a carrier using Bayes' Theorem in formula (1)

where:

A is the outcome (offspring being diabetic or carrier).
B is the given parental genetic history.

P(Carrier | Parental History):

P(Parental History | Carrier) = 1 (since the father is a carrier and the mother is non-diabetic, they can only pass one of the two possible genotypes: DD or Dd).

P(Carrier) = 0.5 (probability of offspring being a carrier from the Punnett Square).

P(Parental History) = 1 (since we are using the known parental genotypes as a given condition).

So,

$$P(\text{Carrier} \mid \text{Parental History}) = \frac{1 \times 0.5}{1} = 0.5$$

P(Non-Diabetic | Parental History):

P(Parental History | Non-Diabetic) = 1 (since the mother is non-diabetic, and the father is a carrier, all offspring will be either DD or Dd, both non-diabetic).
P(Non-Diabetic) = 0.5 (probability of offspring being non-diabetic from the Punnett Square).
P(Parental History) = 1 (as the parent's genetic history is already known).
So,

$$P(\text{Non-Diabetic} \mid \text{Parental History}) = \frac{1 \times 0.5}{1} = 0.5$$

P (Diabetic | Parental History):

P(Parental History | Diabetic) = 0 (since the offspring cannot be diabetic in this case they cannot inherit the dd genotype).

P(Diabetic) = 0 (since it is impossible for the offspring to be diabetic in this case).

P(Parental History) = 1 (same as the previous probabilities, as we know the parental genetic history).

So,

$$P(\text{Diabetic} \mid \text{Parental History}) = \frac{0 \times 0}{1} = 0$$

Final Probabilities:

Using Bayes' Theorem, we calculate the probabilities:

- P(Offspring is Carrier | Parental History) = 0.5
- P(Offspring is Non-Diabetic | Parental History) = 0.5
- P(Offspring is Diabetic | Parental History) = 0

Thus, the probabilities of the offspring's genetic condition, given the parental genetic history, are:

- 50% chance of being non-diabetic (DD).
- 50% chance of being a carrier (Dd).
- 0% chance of being diabetic (dd).

This confirms the earlier conclusion from the Punnett square.

RESULTS AND DISCUSSION

To measure the reliability and validity of the data, we use a confusion matrix. Punnett squares are used to predict the possible genotypes and phenotypes of offspring from a cross between two individuals with known genotypes. Punnett squares visualize possible genetic outcomes in background, while confusion matrices evaluate the performance of classification models in machine learning by comparing predicted and actual values.

To contrast authentic values with predicted values, the machine learning model has a confusion matrix. There are four combinations of predicted and actual values that are given by a confusion matrix: true positive, true negative, false positive, and false negative, which refer to the four types that define the outcome of the prediction exercise. It also works great for determining the accuracy, precision, and recall of a machine learning model.

Since the model is intended to predict future occurrences of diabetes rather than diagnosing current cases, conventional metrics like accuracy may not be directly applicable.

The confusion matrix is derived from a sample dataset of 400 individuals gathered from an endocrinologist maintaining the ethical privacy of the patients, representing a combination of current diabetic and non-diabetic statuses, as well as family history and genetic factors. However, this data is used as a proxy for the genetic predisposition and family history that would affect future generations.

Table 8: Confusion Matrix

	Predicted Diabetic	**Predicted Non-Diabetic**
Actual Diabetic	170 (True Positive)	20 (False Negative)
Actual non-diabetic	20 (False Positive)	163 (True Negative)

This matrix, while based on current data, provides insight into the model's predictive capabilities concerning future generations when genetic predisposition is taken into account.

METRIC CALCULATIONS

The following evaluation metrics were calculated using the confusion matrix, though these should be viewed as proxies for the model's predictive performance rather than an absolute measure of future accuracy.

Accuracy, Precision, Recall and F-Measure, are used for the classification of this work. Table 9 defines accuracy measures below:

Table 9**:** Accuracy Measures

S. No	**Measures**	**Definition**	**Formula**
1	Accuracy	Accuracy is proportion of all classifications that were correct, positive or negative	$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$
2	Precision	Precision is proportion of all the positive classifications that are actually positive	$Precision = \frac{TP}{TP + FP}$
3	Recall	Recall refers to the proportion of all actual positives that were classified correctly as positives	$Recall = \frac{TP}{TP + FN}$
4	F-Measure	F-Measure is the weighted average of both precision and recall	$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$

From the above-mentioned confusion table accuracy, precision, recall and F1 score can be calculated:

1. **Accuracy**:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Accuracy = \frac{170 + 163}{170 + 163 + 20 + 20}$$

$$Accuracy = \frac{333}{400}$$

$$Accuracy = 83.2\%$$

2. **Precision:**

$$Precision = \frac{TP}{TP + FP}$$

$$Precision = \frac{170}{170 + 20}$$

$$Precision = \frac{170}{190}$$

$$Precision = 89.5\%$$

3. **Recall:**

$$Recall = \frac{TP}{TP + FN}$$

$$Recall = \frac{170}{170 + 20}$$

$$Recall = \frac{170}{190}$$

$$Recall = 89.5\%$$

4. **F1 Score:**

$$F1\,Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

$$F1\,Score = 2 \times \frac{0.895 \times 0.895}{0.895 + 0.895}$$

$$F1\,Score = 89.5\%$$

Stability of the proposed predictive model is illustrated with a confusion matrix that shows the model's performance in differential diagnosis of diabetes. Within a sample of 400 cases from reality, the model achieved a Diabetic Detection Rate of 170 (True Positives) and a Non-Diabetic Detection Rate of 163 (True Negatives). In addition, there were 20 misclassifications of Diabetics as Non-Diabetics (False Negatives) and 20 Non-Diabetic individuals classified as Diabetics (False Positives). These values demonstrate high levels of accuracy and reliability with only a few misclassifications. The derived metrics provided an accuracy of 89.3%, precision of 89.5%, alongside a strong recall rate, underscoring stable performance regardless of the scenario. Such stability validates system predictive stability as the system is able to repeatedly provide accurate results across rigorously and ethically gathered diversified family genomic data.

CONCLUSION

This research demonstrates the potential of combining classical Mendelian genetics with probabilistic modeling to predict the risk of diabetes to future generations. By integrating the Punnett square methodology with the Naïve Bayes algorithm, the proposed system offers an intuitive tool to forecast hereditary diabetics.

This application was practically tested on real-world data taken from 400 families, recording an accuracy of about 83.2 percent, having high precision and recall; thus, confirming the effectiveness of this model. Importantly, the model also testified the importance of carrier status in genetic inheritance, even in non-diabetic individuals. This understanding is critical for preventive health approaches in countries such as Pakistan, where diabetes prevalence is rapidly increasing.

In short, in this research contribution the Mendelian Genetics provides deterministic, rule-based predictions. While Naïve Bayes Algorithm use for statistically bounded, repeatable probabilistic outcomes included desire results. Confusion matrix confirms consistency and performance of the model.

Moving forward, this system can be extended further to include other genetically influenced conditions such as kidney disease, cardiovascular issues, and eye disorders.

ACKNOWLEDGEMENTS

I, Syeda Azwa Asif as the corresponding and the first author of the paper acknowledge the contribution and efforts of the co-authors for their support and interest in the topic till the completion of the work.

CONFLICT OF INTEREST

The authors are completely truthful in their claim that they have no conflicting interests with regard to this work. Every single viewpoint and the conclusion made in this research belong to the

authors alone and are entirely unaffected by any financial, personal, or professional forces that might be assumed to be affecting the objectivity of this research.

ETHICAL INFORMATION

Current study derived or analyzed datasets are not attainable online due to ethical and privacy restrictions, but could be availed from corresponding author Syeda Azwa Asif upon reasonable request. The data was collected with the cooperation of endocrinologists and is subjected to confidentiality agreements to protect the privacy of individuals involved.

FUNDING

I, the corresponding author and my coauthors would like to declare that this project was completed without any special financial support or funding from any organizations. All obtained resources were either self-funded or voluntarily contributed by the research group members themselves. In the future we seek possible collaborations or assistance to extend this work for which we will welcome and appreciate the investors.

AUTHOR CONTRIBUTION

Syeda Azwa Asif, the primary author, initialized the main research idea, developed the strategy and methodology for implementing and conceptualizing the project, and played the lead role in developing the intelligent software application. She led the study design, collected and analyzed the data, and also managed the revisions and overall coordination of the research with the co-authors.

Syed Asif Ali, provided expertise in artificial intelligence and machine learning algorithms. Also contributed to refining the study design, supervised the application of Naïve Bayes, and reviewed the manuscript critically for important intellectual content.

Asma Khan provided expertise in coding, development and contributed to the review. She also actively participated in data analysis and mapping design.

REFERENCES

Admin. (2023, January 5). *Mendel's Laws of Inheritance - Mendel's Laws and Experiments*. BYJUS. https://byjus.com/biology/mendel-laws-of-inheritance/

Bateson, W., & Mendel, G. (2013). *Mendel's principles of heredity*. Courier Corporation.

Berrar, D. (2019). Bayes' theorem and naive Bayes classifier.

Birjais, R., Mourya, A. K., Chauhan, R., & Kaur, H. (2019). Prediction and diagnosis of future diabetes risk: a machine learning approach. *SN Applied Sciences*, *1*, 1-8.

Davis, L. C. (1993). Origin of the Punnett square. The American Biology Teacher, 55(4), 209-212.

Edwards, A. W. F. (2012). Punnett's square. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *43*(1), 219-224.

Gautier, T., Ziegler, L. B., Gerber, M. S., Campos-Náñez, E., & Patek, S. D. (2021). Artificial intelligence and diabetes technology: a review. *Metabolism*, *124*, 154872

Genetics, autosomal dominant. (2024, January 1). PubMed.

Germain, D. P., & Jurca-Simina, I. E. (2018). Principles of human genetics and Mendelian inheritance. Neurometabolic hereditary diseases of adults, 1-28.

International Diabetes Federation. (2023, June 7). Pakistan - International Diabetes Federation.

Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. *Computational and structural biotechnology journal*, *15*, 104-116.

Mujumdar, A., & Vaidehi, V. (2019). Diabetes prediction using machine learning algorithms. *Procedia Computer Science*, *165*, 292-299.

Müller-Wille, S., & Parolini, G. (2020). Punnett squares and hybrid crosses: how Mendelians learned their trade by the book. BJHS Themes, 5, 149–165.

Pakistan diabetes report 2000 — 2045. (n.d.). International Diabetes Federation.

Prasad, R. B., & Groop, L. (2015). Genetics of type 2 diabetes—pitfalls and possibilities. Genes, 6(1), 87-123.

Priya, K. L., Kypa, M. S. C. R., Reddy, M. M. S., & Reddy, G. R. M. (2020, June). A novel approach to predict diabetes by using Naive Bayes classifier. In *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184)* (pp. 603-607). IEEE.

Rossi, A., & Kontarakis, Z. (2022). Beyond Mendelian inheritance: Genetic buffering and phenotype variability. Phenomics, 2(2), 79-87.

Sebastiani, P., Solovieff, N., & Sun, J. X. (2012). Naïve Bayesian classifier and genetic risk score for genetic risk prediction of a categorical trait: not so

Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., &

different after all!. *Frontiers in genetics*, *3*, 26.

Sisodia, D., & Sisodia, D. S. (2018). Prediction of diabetes using classification algorithms. *Procedia computer science*, *132*, 1578-1585.

Tapak, L., Mahjub, H., Hamidi, O., & Poorolajal, J. (2013). Real-data comparison of data mining methods in prediction of diabetes in Iran. *Healthcare informatics research*, *19*(3), 177-185.

World Health Organization: WHO & World Health Organization: WHO. (2023, April 5). Diabetes.

Xu, S. (2022). Review of Mendelian Genetics. In Quantitative Genetics (pp. 13-24). Cham: Springer International Publishing.

Yu, W., Liu, T., Valdez, R., Gwinn, M., & Khoury, M. J. (2010). Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes. *BMC medical informatics and decision making*, *10*, 1-7.

Yule, G. U. (1902). Mendel's laws and their probable relations to intra-racial heredity (continued). *New Phytologist*, *1*(10), 222-238.

Zschocke, J., Byers, P. H., & Wilkie, A. O. (2023). Mendelian inheritance revisited: dominance and recessiveness in medical genetics. Nature Reviews Genetics, 24(7), 442-463.