HIGH SCHOOL EDITION
Journal of Student Research

Volume 11 Issue 3 (2022)

Mathematical Ranking of AI Classifiers Using Confusion Matrix and Matthews Correlation Coefficient

Anshuman Bhamidipati

Conant High School, Hoffman Estates, IL, USA

ABSTRACT

This research primarily deals with mathematically ranking the level of accuracy of various Artificial Intelligence (AI) based machine learning classifiers, using Mathews Correlation Coefficient (MCC), leveraging Confusion Matrices. A detailed Literature survey was done to gather the existing knowledge. This knowledge was used as foundational basis to further build the scope of this research project. The classifiers used were Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Convolutional Neural Network (CNN). These classifiers were used for Face recognition of individuals. A total of 33 adult test subjects with 10 images per subject resulting in a total of 330 distinct images were part of this research project. These test subjects were from the *Labeled Faces in The Wild* publicly available database along with teachers from James B. Conant High School who voluntarily participated in this research. Python based face recognition program wrappers and associated environment was built around pre-existing classifiers and the image data was passed to these wrappers. These wrappers recognized the faces of individuals in the images over 10 trials wherein each trial consisted of 33 distinct images, with varying degree of accuracy and presented that as outputs. These outputs were used to determine the Confusion Matrices which in turn were used to calculate the MCC scores. The MCC scores were plotted, and these results showed that the SVM AI classifier had the highest level of relative accuracy for face recognition, followed by KNN and CNN classifiers.

Introduction

Artificial Intelligence (AI) is the development and implementation of computer systems that can perform tasks that normally require human intelligence on their own. These systems are characterized by their ability to draw patterns from data that they are presented with (Patrick & Williams, 2020). There are two categories of AI, Artificial General Intelligence (AGI) and Artificial Narrow Intelligence (ANI). AGI, is characterized by computer systems that have the ability of human level thought and perception of the world. One of the main misconceptions related to AI, is that it is thought by many people that current level of AI technology has attained AGI. But this is completely incorrect as, modern day AI technology can only perform very narrow tasks such as detecting obstacles in a path, recognizing speech, detecting written characters, and more. This kind of narrow intelligence associated to a single task completed by AI, is known as ANI.

Artificial Intelligence is a broad umbrella under which there is a more specific category known as Machine Learning (ML). ML is the field of study that gives computers the ability to learn without being explicitly programmed. Under the umbrella of ML, there is a narrower field known as Deep Learning which is essentially the same thing as an Artificial Neural Network. This type of ML implementation only requires an input and desired output, in the form of data. The software builds the internal neurons and synapses on its own, after specifications are defined. Neurons and synapses are a specialized way in which Artificial Neural Networks process the data that they are given, to produce meaningful results. The data is pushed through each layer of the neurons with forward propagation. The specialty of these neurons and synapses are that they can be reconfigured based on the accuracy of the results through back propagation, allowing the Artificial Neural Network to learn from its previous mistakes.

One of the specific uses of AI and ML, which will be discussed in this paper, is the field of Computer Vision. Computer Vision is a field of science that deals with how computers can gain high-level understanding from digital images or videos. More specifically, this paper will deal with the usage of Computer Vision to perform Face Recognition. Face Recognition is a method of identifying or verifying the identity of a person with the use of their face in the form of a picture, video, or in real-time.

Literature Review

Within the field of Computer Vision, Face Recognition can be achieved through various methods, that yield different levels of success, mainly in the form of accuracy in recognizing a human face. A face recognition classifier is a unique model and the underlying base algorithm that employs the usage of different theories to process data in a specialized way to gain meaningful results from it. Most of modern research in this field of AI is in perfecting and developing new classifiers or unique classifier-based programs. In the article, *Optimization of K-nearest neighbor using particle swarm optimization for face recognition*, the authors developed a novel approach to recognize human faces, with the K-nearest neighbor classifier (KNN). The authors used the combination of classifiers and methodologies to build a program that can recognize human faces and verify them. The authors tested their novel approach by collecting data from 155 subjects along with data from ORL dataset and found that their model had a relatively accurate system, as they declared at the end of their paper (Sasirekha & Thangavel, 2019).

As AI technology grew in popularity and developed to have higher levels of accuracy and efficiency, some ethical concerns were raised which are relevant to be discussed here. In the paper, *Should We Trust Artificial Intelligence?*, it was presented that with any AI system comes the key aspect of being able to trust it and put one's faith into it, so that its benefits can be fully reaped. The main idea being presented was that trust on AI systems does not stem from the system itself, but rather from the person, institution, or company that built the AI system. And that the AI implementation can only be trusted when its creators are trustworthy (Sutrop, 2019). This is completely true, and there are guidelines that have already been put into place so that AI development can be accomplished safely and ethically without any discrimination against any individuals' physical characteristics.

While AI based Face Recognition, is not widely implemented, and integrated into the education system, other AI based systems have been implemented. Several articles including, *Artificial Intelligence and the Student Experience: An Institutional Perspective*, have discussed that AI is beneficial to schools and the students. Current technology of AI has been implemented in the form of chatbots that have helped guide students and provide learning supports, automated paper grading systems, and more (Khare et al., 2018). These systems have greatly benefited students by giving them access to lots of relevant information and feedback in a short period of time. And has allowed counselors, teachers, students service members, and other school staff to further education with the increased amount of time avaliable.

Looking at the abilities of AI systems, it is imperative for some to have a sense of fear that teaching would be overtaken by intelligent programs and robots. But that is not entirely true as was presented in the research article, *Can Modern AI replace teachers? Not so fact! Artificial Intelligence and Adaptive Learning: Personalized Education in the AI age*, where different amounts of data generation from different students proved that current AI applications cannot replace experienced teachers, as they would only benefit students who generate larger amounts of data, because AI systems determine patterns out of data and based on these patterns, help the students. AI systems and applications are at a level of development, that instead of replacing teachers, they can be used as teacher assistants (Kolchenko, 2018). Through the above two mentioned articles, AI applications are highly beneficial for their implementation in the education system, but experienced teachers are still necessary for the advancement of knowledge.

The Knowledge gap where my research falls is that many classifiers that have been developed for face recognition, are not correlated with each other from an accuracy standpoint. This research aims to compare some of these classifiers using face images, from datasets and volunteered teachers, in order to find the relative accuracy through the use of a Confusion Matrix. This confusion matrix is used to determine Matthews Correlation Coefficient (MCC), so that the relatively best classifier can be found. The MCC, not only provides information on the AI Face Recognition system's accuracy but it also gives deeper analysis such as True Positives, True Negatives, False Positives, and False Negatives, which will be further explained in the methods section of this paper. The hypothesis of this research is that AI classifiers that can derive a higher amount of data from images would perform with a higher relative accuracy compared to other classifiers.

Method

The methodology that was employed in this research, closely resembled the quasi-experimental alternating treatment research design, but had some changes. Traditionally the quasi-experimental alternating treatment research design asks for recording several observations over the course of a time period, while making multiple interventions to change the independent variable, in order to observe the changes of a single dependent variable. But, in my research design instead of a single dependent variable, three different dependent variables were used.

The quasi-experimental alternating treatment research design was the most logical and accurate research design for the purpose and intent of this research. This is because the mentioned research design allowed me to correctly assess the accuracy of the 3 dependent variables, which were different classifier-based Python (programming language) programmed wrappers designed to accomplish facial recognition, while being able to provide a changing independent variable (images that were used as inputs). Other research design methodologies such as conducting a survey can be used to collect the viewpoints of a huge number of individuals on this topic of face recognition, but it cannot fulfill the purpose of my research project to rank the classifiers. Same way, another design methodology of conducting detailed, in-depth interviews, to gather the viewpoints from individuals does not help as it cannot provide the accuracies of face recognition classifiers. A case study-based approach was also originally considered, but later discarded because it was much more deeply focused on gaining qualitative information, whereas quantitative data was needed for greater understanding of the topic. Mainly, the discarded research designs would not provide me the ability to test the live performance of different classifier based facial recognition programs. Due to this reason, all but the experimental research designs were eliminated. Since multiple image sets would need to be passed to the face recognition programs in order to truly test their performance, the alternating treatment research design was narrowed upon and implemented (Leedy et al., 2019, pp. 205 - 211).

The next step of the experiment was to successfully build the AI environment. Python (programming language) coding was done to connect different code blocks together. Python face recognition programs based on different classifiers were used as blocks. The PyCharm Python IDE (Integrated Development Environment) was leveraged for coding.

The first program utilized a Support Vector Machine (SVM) classifier. SVM classifiers work by building a hyperplane between sets of data, and then using this hyperplane to make predictions for the face recognition of individuals in the testing images. The SVM classifier was implemented using OpenCV and Dlib libraries within Python (Hassan, 2020). A library in programming is a collection of various functions that can be utilized to help perform different tasks.

The second program utilized a K-Nearest Neighbors (KNN) classifier. A KNN classifier works on a single main assumption that similar data will exist in close proximity to each other. With this assumption a KNN classifier, in the case of face recognition, groups the data for different individuals so that when the testing images are provided, similar data points derived from the images are recognized as similar individuals. The above described KNN classifier was implemented using the Scikit-Learn library (*ML | Implement*, 2019).

The third program utilized a Deep Convolutional Neural Network (CNN) classifier. A CNN consists of an input layer, hidden layers, an output layer, neurons, and synapses. The neurons and synapses are used to compute different data points of the face images, that are defined by the computer and each set of these neurons and synapses is a single layer. By repeating this process of using neurons and synapses, multiple layers are built, each computing different data points from the images. Through the process of forward propagation data sets derived from the images are passed through all the layers, to finally reach the output layer, where the individual in the image is predicted. This CNN classifier was implemented in the program using the TensorFlow library with Keras backend (Skúli, 2018).

Then several lines of code were added to these face recognition programs in order to allow them to work for my application, my interface, the image sets being passed, and for displaying the results back to me.

After the programs were built, the images that were required to be passed to the programs needed to be sourced. Originally it was planned for students to be included as participants in the experiment after distribution of consent forms, but due to privacy restrictions and school policy restrictions this was not possible. So, as a revised plan a total of 33 adult participants were part of this research project, both from a publicly available database and teachers from James B. Conant High School. The publicly available database used is known as *Labeled Faces in The Wild* database. This contained color images of famous adult individuals. In addition to the database, teachers participated in this research project, after they signed and agreed to a consent form [Appendix B] that was distributed to them. The set of participants involved in this research project were chosen to be demographically vivid, as this is a key factor for the performance of face recognition programs. Overall, 10 color images of each participant with different facial expressions were used, which equaled a total of 330 images for 33 adults, that were then randomly sorted into 10 image sets.

After the images were sourced and formatted, they were organized into a Database set and Testing sets. The Database set, to train the classifiers, was intended to have a single image per person in the database. But, for 33 individuals, intentionally only 21 individuals were populated, in order to build the confusion matrix. The Testing sets, to test the accuracy of classifiers, comprised of all 330 images grouped into 10 sets, with one image per individual. All these images were passed to the three classifier programs. After the programs analyzed the images that were dedicated as database images, the programs moved on to recognize faces in the 10 testing image sets. Once the programs finished executing, I was able to see the faces that were and were not recognized by the programs. In order to better analyze these results, an Excel spreadsheet model was built with MCC, precision, and recall formulas in it. From the results of the 3 programs the spreadsheet model calculates the confusion matrices, which organizes the statistics and type of the data, and were necessary for the next step of using the Matthews Correlation Coefficient (MCC).

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Equation 1. Formula of Matthews Correlation Coefficient

Equation 1 shows the MCC equation that was utilized, wherein True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN) values are used. The advantage of using MCC, was the handling of irregular datasets as opposed to other correlation models which cannot work as effectively with irregular datasets. The MCC score is a value between -1 and 1, where 1 is a perfectly functioning program and a -1 is a completely incorrect program.

The Excel spreadsheet model developed by me, accurately and consistently determined the MCC score of the three programs based on the TP, TN, FP, and FN values used in the confusion matrix. These values were then used to rank the three programs relative to each other, determine the best program, and to further determine the margin of accuracy between the programs results.

Results and Findings

The confusion matrices calculated, for SVM, KNN, and CNN classifiers in my experiment are tabulated in Appendices F, G and H. This can be used to isolate the True Positives, True Negatives, False Positives, and False Negatives from the different images that were recognized and were not recognized by the system.

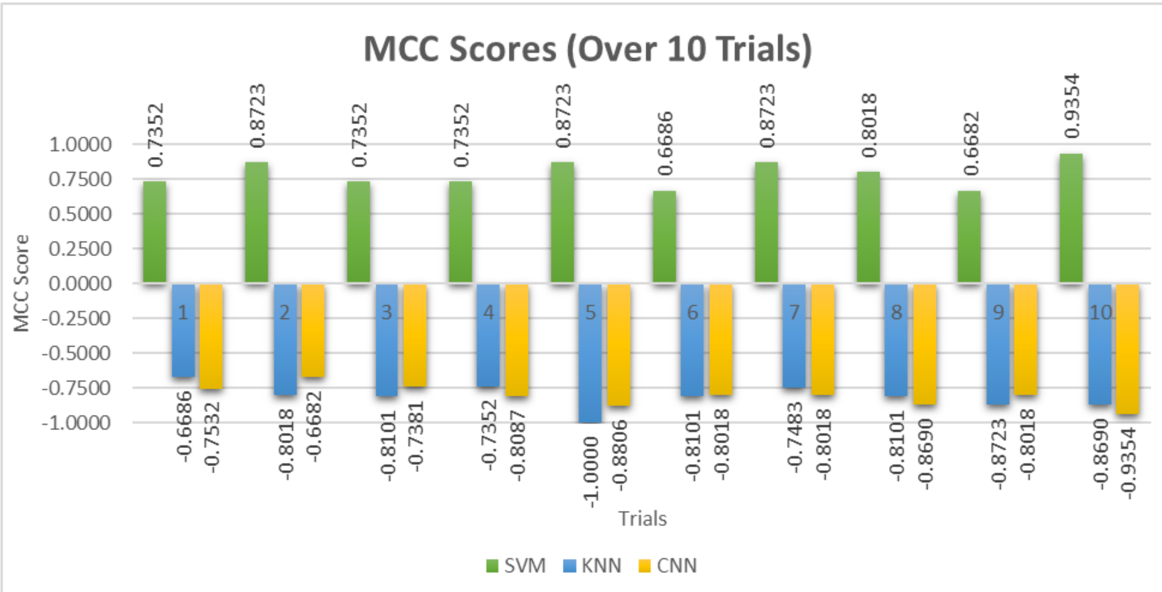

Figure 1. MCC Scores of SVM, KNN, and CNN classifier-based models over the 10 trials.

Figure 1 shows the distribution of the MCC scores of the SVM, KNN, and CNN classifier-based models over the 10 trials. The MCC, measures the accuracy of the model within a range of -1 to 1, where a score of 1 signifies a perfect system, a score of 0 signifies a system equivalent to making random choices, and a -1 signifies a system that makes incorrect choices all the time. The SVM classifier has the highest MCC score out of all the classifiers and is fairly accurate, as it is consistently above the MCC score of 0.6000. The CNN classifier has the second highest MCC score because, while it may seem to be equal to the KNN classifier, an average of the individual MCC scores over the ten trials is -0.8059. This average is higher than the average of the ten individual MCC scores for the KNN classifier, which is -0.8126, allowing KNN to have the third highest MCC score.

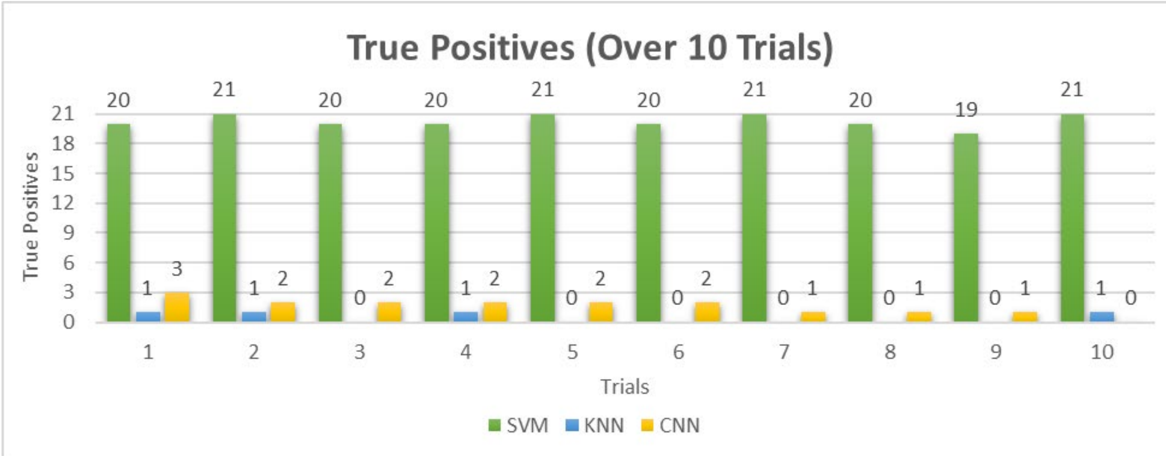

Figure 2. True Positive values of SVM, KNN, and CNN classifier-based models over the 10 trials.

In the context of my project the true positive data represents the test subjects, whose face image was provided as a database image to the model and was correctly recognized by the model in the testing images. Figure 2 shows the true positive values of the SVM, KNN, and CNN classifier-based models over the course of the 10 trials. Overall, the SVM classifier-based model has the highest true positive values, which are ideal in this situation. Followed by the CNN classifier-based model which has the second highest true positive values. And finally, the KNN classifier-based model has the third highest true positive values.

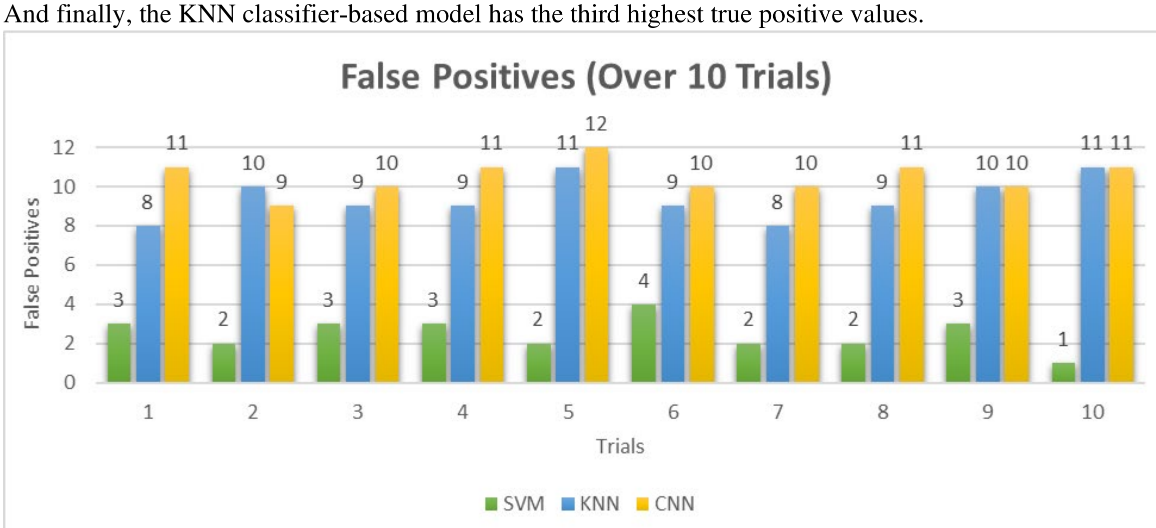

Figure 3. False Positive values of SVM, KNN, and CNN classifier-based models over the 10 trials.

The false positive data represents the test subjects, whose face image was not provided as a database image to the model, but the model still recognized them as someone they were not. Figure 3 shows the false positive values of the SVM, KNN, and CNN classifier-based models over the ten trials. Comparatively, the SVM classifier-based model has the lowest overall false positive values, which is ideal for an accurate AI system. By comparing all the values over all ten trials, the KNN classifier-based model has the second lowest false positive values, and the CNN classifier-based model has the third lowest false positive values.

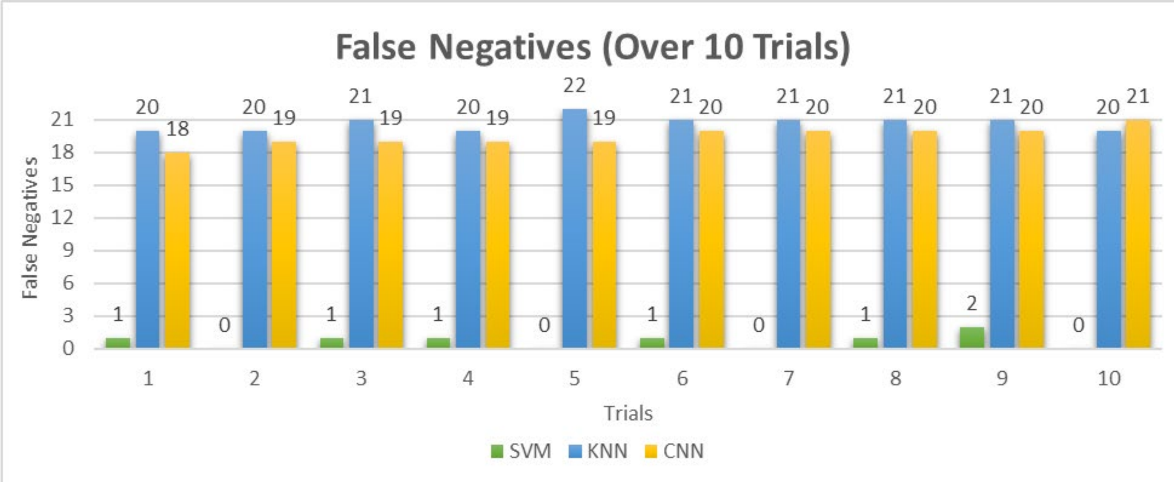

Figure 4. False Negative values of SVM, KNN, and CNN classifier-based models over the 10 trials.

The false negative data represents the test subjects, whose face image was provided as a database image to the model and was either incorrectly recognized or not recognized at all. In Figure 4, it can be seen that, the SVM classifier-based model has the lowest overall false negative values, which is ideal for an accurate face recognition AI system. Through the comparison of all the values over the 10 trials, it can be seen that the CNN classifier-based model has the second lowest false negative values and the KNN classifier-based model has the third lowest false negative values.

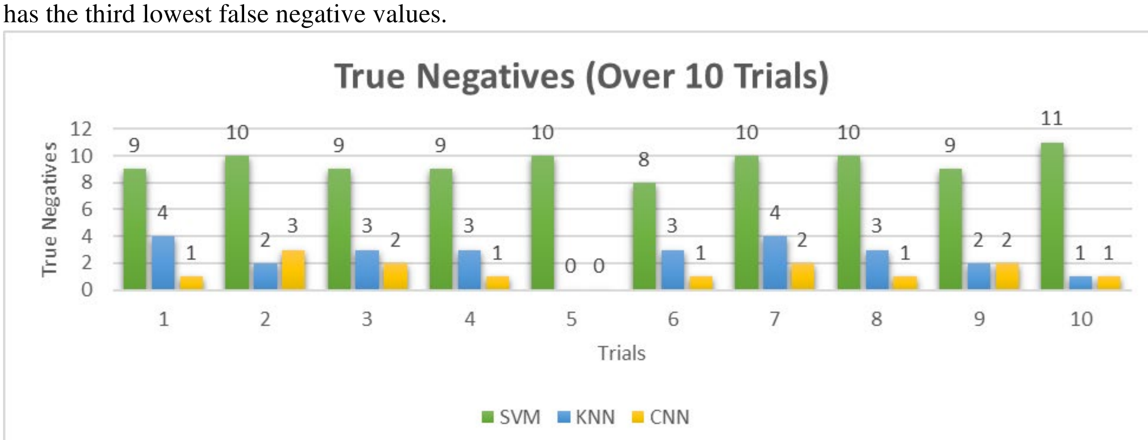

Figure 5. True Negative values of SVM, KNN, and CNN classifier-based models over the 10 trials.

The true negative data represents the test subjects, whose face image was not provided as a database image to the model and the model didn't recognize the face. Throughout all the trials visualized in Figure 5, the SVM classifier-based model has the highest true negative values, which is the goal of any accurate face recognition AI system. The KNN classifier-based model has the second highest true negative values, and the CNN classifier-based model has the third highest true negative values.

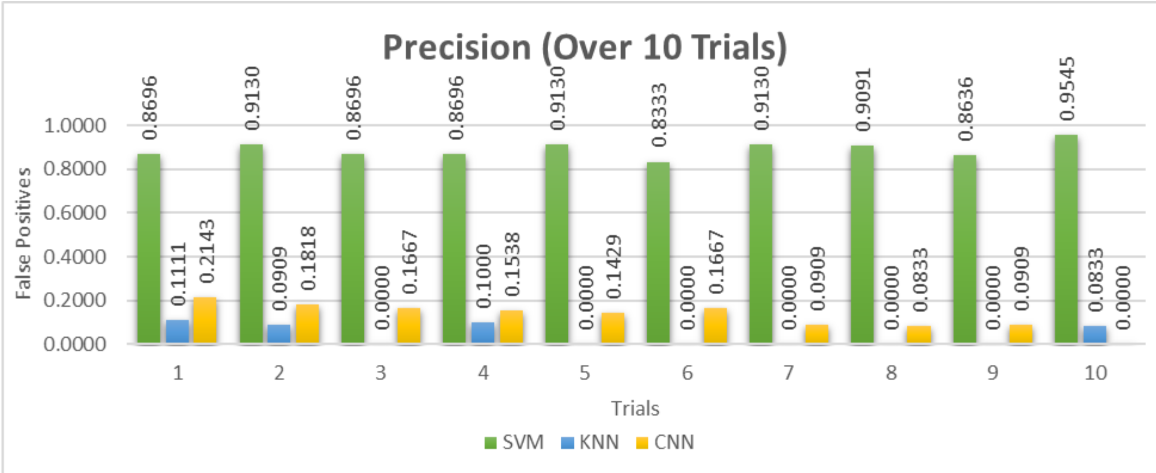

Figure 6. Precision values of SVM, KNN, and CNN classifier-based models over the 10 trials.

In the context of my project, the precision values show the fraction of all the images that were classified correctly among all the images that were classified by the programs. The precision of the systems, as displayed in Figure 6, can have a theoretical maximum of 1 and theoretical minimum of 0. As can be seen in the graph, the SVM classifier-based model has the highest precision, followed by the CNN classifier-based model which has the second highest precision, and finally the KNN classifier-based model with the third highest precision values.

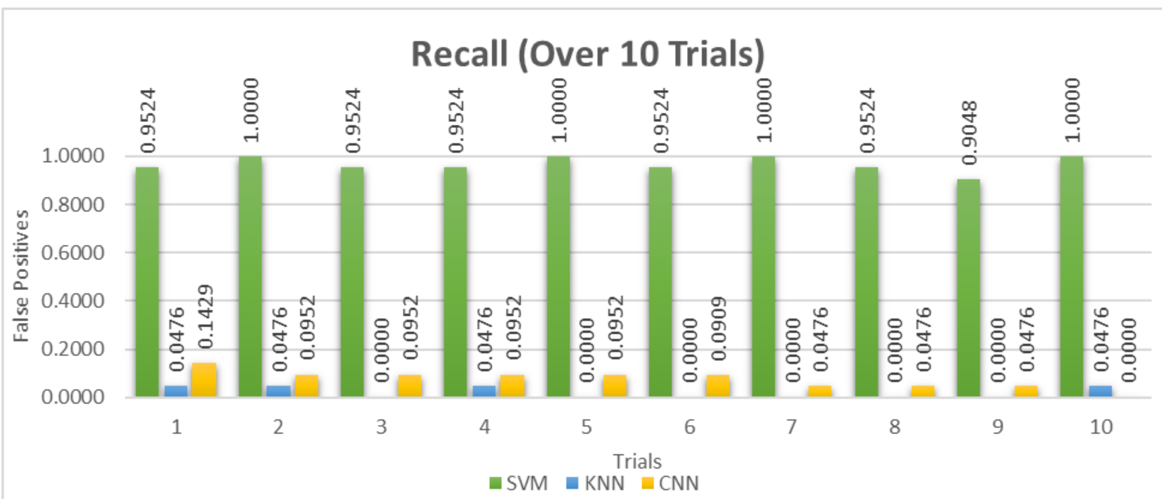

Figure 7. Recall values of SVM, KNN, and CNN classifier-based models over the 10 trials.

In the context of my project, the recall values show the fraction of all the images that were classified correctly among all the images that were present in the database of the programs. Similar to the precision values, recall values shown in Figure 7 can have a theoretical maximum of 1 and a theoretical minimum of 0. Throughout the 10 trials, the SVM classifier-based model has the highest recall values, followed by the CNN classifier-based model, and closely followed by the KNN classifier-based model with the third highest recall values.

Discussion and Analysis

Regular non-AI based programs need to be given specific instructions and commands on what needs to be done to understand a face or an image, whereas a machine learning (ML) based AI program need not be given specific instructions because these programs learn by themselves. Hence using an AI based Face Recognition program is very advantageous. Face Recognition is a two-step process, first the programs involved need to get trained with some data and then the same set of programs will use the training knowledge obtained to identify the faces given to them. The same approach was followed in this project, for SVM, KNN, and CNN classifiers.

The SVM program, reads the faces in the database images and based on 128 points on each face, and distances between those points, identifies a face. This 128 point approach is the specialty of SVM unlike KNN and CNN. In SVM, the classifier does not need to be told where to mark those 128 points on face as it does this step all by itself. Next, the SVM program determines the key 128 points on test image and compares that with points and distances already collected from the first run (database image), and then cycles through all the database images, through the use of a hyperplane. This hyperplane allows the program to find the smallest possible difference between the database and testing images to generate a similarity percentage and identify the person in the test image. This approach gives SVM an edge over other classifiers.

Whereas the KNN classifier bases its decision on the nearest neighbors of the testing images data. Initially the KNN classifier, in this project, determines the face data for all the faces in the database and plots them on a cartesian coordinate plane. Then the KNN classifier determines the face data for each of the faces in the testing images and then plots them on the same cartesian coordinate plane. It then searches for five nearest neighboring points in the coordinate plane from the testing images data and determines the face of the person based on the number of closest data points that have the same face data. So, from the MCC scores of KNN classifier (Figure 1) it can be seen that this neighbor-based approach followed by KNN is not great as compared to SVM.

The CNN classifier-based program is a little different from the above mentioned classifiers, as it determines and calculates the face data based on neurons and synapses approach, through forward and backward data propagation to find the closest match. But based on MCC scores (Figure 1), this approach is again not great as compared to SVM classifier.

Originally, my hypothesis stated that the more data a program collects, the higher are its chances of providing the correct identification. But, from the experiment it can be seen that this is partially true. While it is true that a higher amount of data can lead a ML based program towards correct face identification, the internal mechanism and workings of the classifiers also plays a vital role on the accuracy of the program, as is evident from the MCC data produced by SVM classifier.

Secondly, this research, unlike many others (Sasirekha & Thangavel, 2019), tested the accuracy and effectiveness of ML based AI classifiers when they only have access to a single image per individual in the database. It has been found that classifiers have a higher accuracy when they have access to multiple images of a single individual in the database. This was another reason that led me to make my hypothesis. But, in places such as schools, which is one of the driving factors in my research, it would be a better alternative to have less images per individual. Due to this fact, my project focused on single images per individual. Surprisingly, though the SVM classifier-based program was still able to achieve a high accuracy rate as shown in all the data collected (Figure 1 - 7). In all the measurements, the SVM classifier was closest to the ideal, compared to all the other classifiers. This astounding result must be primarily since the SVM classifier-based program compares a single testing image to a single database image to determine a match. And due to this process, all demographic based errors were eliminated, and the confusion of the program when referencing multiple images at once is also eliminated, allowing for targeted analysis and accurate prediction.

Implications

This research has many implications to schools and to the research on AI, ML, and Face Recognition that is already present.

With this project and the data collected, it can be clearly seen that the SVM classifier-based program was able to achieve the highest accuracy compared to the other programs. This can be vital information specifically for high schools that are planning on exploring the possibility of incorporating AI and Face Recognition into their processes. As this shows that schools can utilize the SVM classifier-based program and have trust that it functions to a very high level of accuracy. After testing the SVM classifier-based program on the images from the *Labeled Faces in The Wild* database, it can be said that it works on other faces including teenagers and staff members present in high schools. This program also successfully recognized faces of multiple different demographics, which can be another reassurance, if schools choose to consider it as a component in modern surveillance devices.

This research would also add extra data and results to the existing and growing field of AI and Face Recognition, by giving information on the performance of classifiers with the use of only one image per individual in the database. This added information in the wider body of research can be useful in developing more sophisticated, powerful, and mainly accurate Face Recognition systems.

Limitations

One of the limitations of this research was that student faces were not used to test the performance of the classifier-based Face Recognition programs, instead faces of individuals from the *Labeled Faces in The Wild* database and teachers were used to access the performance of the programs. Due to this, no direct correlation can be confidently established between the accuracy of classifier-based programs, to their possible effectiveness on student faces (as the age group of students is less than the teachers and individuals in the database). But, due to the extremely high accuracy of the SVM classifier-based program, it can be stated that it would be effective against students faces as well. Another, limitation of my research project was that due to the Covid-19 pandemic, I was not able to visit my school and collect video data of teachers faces, which would have been used as another test scenario.

Conclusion

In the beginning of this research, I hypothesized that Face Recognition programs that collect a larger quantity of data would be more accurate in their predictions. But, upon the end of this research, it was found that the quantity of the data collected combined with the internal mechanism and workings of the classifiers yielded a good prediction by the program. Through this research it was found that the SVM classifier-based program was the most accurate in all aspects of measurement, compared to all the other programs that were used in this research. But a significant detail of this research project that cannot be overlooked is that these programs were tested on their accuracy while only having one image for every individual in their database. This is unlike many other researches, for example (Sasirekha & Thangavel, 2019), where multiple images were used in the database for every individual. The data collected and results derived from this research can directly expand the existing knowledge and data related to ML based AI Face Recognition.

This technology of Face Recognition is highly useful, and the possibilities of its application would grow exponentially, by having only a single image per individual in the programs' database, as the SVM classifier had, in this research. An example of this usage would be in schools that would have many benefits. By only having to obtain a single image per student, the school administration can easily save time, cost, and resources. Face Recognition can also be used in many other places such as libraries, supermarkets, workplaces, transit services, and many more. This technology can be used across the globe, for surveillance purposes.

Future Directions

This research can be taken forward and expanded upon to collect more data and derive deeper results. Specifically, this may include, increasing the number of distinct classifier-based programs being tested such as Random Forest, Adaboost, etc., while retaining the measuring scale of the MCC, for more accurate comparison and analysis of results. Also, a greater quantity of images can be used to test these programs. To collect greater and valuable data, the types of images used should continue to be diverse and varied. Examples of these variations include the light levels of the images, orientation of the face in the image, increasing the number of demographics of people that are tested, varying the age range of people, varying the pixel quantity of images, and many more. If accessible, these programs may be connected to GPU's (Graphics Processing Units) and TPU's (Tensor Processing Units) to greatly decrease the time it takes to process images. With this a greater quantity of images can be tested in a shorter amount of time, thus increasing the data collected for each individual program. Due to technological limitations, GPU's and TPU's were not accessible in this research, so data collection was a more time-consuming process.

These steps to carry forward this research are highly useful, as they can lead to multiple benefits. One of these possible benefits, may be the finding of a specific classifier-based program that has a higher accuracy, and therefore is more suitable for real world application. Another monumental benefit is that, with a greater quantity of data being collected, more relationships between the factors of Face Recognition programs can be uncovered. These new results and relationships can then lead to the breakthroughs in development of even more accurate and efficient Face Recognition systems.

Acknowledgments

I would like to thank my advisor for the valuable insight provided to me on this topic.

Works Cited

Khare, K., Stewart, B., & Khare, A. (2018). Artificial Intelligence and the Student Experience: An Institutional Perspective. *IAFOR Journal of Education*.

Kolchenko, V. (2018). Can Modern AI Replace Teachers? Not so Fast! Artificial Intelligence and Adaptive Learning: Personalized Education in the AI Age. *HAPS Educator*, 249-252.

Patrick, B., & Williams, K. L. (2020). What is artificial intelligence? *Journal of Accountancy*, *229*(2), 1-4.

Sasirekha, K., & Thangavel, K. (2019). Optimization of K-nearest neighbor using particle swarm optimization for face recognition. *Neural Computing & Applications*, *31*(11).

Sutrop, M. (2019). SHOULD WE TRUST ARTIFICIAL INTELLIGENCE? *TRAMES: A Journal of the Humanities & Social Sciences*, *23*(4), 499-522.

Hassan, M. (2020, June 11). *Face Recognition + Attendance Project | OpenCV Python (2020)* [Video]. YouTube. https://youtu.be/sz25xxF_AVE

Leedy, P. D., Ormrod, J. E., & Johnson, L. R. (2019). *Practical research: Planning and design* (12th ed.). Pearson Education.

ML | Implement Face recognition using k-NN with scikit-learn. (2019, March 15). GeeksforGeeks. Retrieved December 29, 2020, from https://www.geeksforgeeks.org/ml-implement-face-recognition-using-k-nn-with-scikit-learn/

Skúli, S. (2018, January 11). *Making your own Face Recognition System*. freeCodeCamp. Retrieved December 31, 2020, from https://www.freecodecamp.org/news/making-your-own-face-recognition-system-29a8e728107c/

Works Consulted

Ahmad, Z., & Shahzadi, E. (2018). Prediction of Students' Academic Performance using Artificial Neural Network. *Bulletin of Education and Research*, *40*, 157-164.

Barrett, M., Branson, L., Carter, S., DeLeon, F., Ellis, J., Gundlach, C., & Lee, D. (2019). Using Artificial Intelligence to Enhance Educational Opportunities and Student Services in Higher Education. *Inquiry*.

BATTINI SÖNMEZ, E. (2019). A computational study on aging effect for facial expression recognition. *Turkish Journal of Electrical Engineering & Computer Sciences*, *27*(4).

Beaulac, C., & Rosenthal, J. S. (2019). Predicting University Students' Academic Success and Major Using Random Forests. *Research in Higher Education*, *60*(7).

Chomanski, B. (2019). What's Wrong with Designing People to Serve? *Ethical Theory & Moral Practice*, *22*(4), 993-1015.

González-Hernández, F., Zatarain-Cabada, R., Barrón-Estrada, M. L., & Rodríguez-Rangel, H. (2018). Recognition of learning-centered emotions using a convolutional neural network. *Journal of Intelligent & Fuzzy Systems*, *34*(5).

Hinojo-Lucena, F.-J., Aznar-Díaz, I., Cáceres-Reche, M.-I., & Romero-Rodríguez, J.-A. (2019). Artificial Intelligence in Higher Education: A Bibliometric Study on Its Impact in the Scientific Literature. *Education Sciences*.

Huang, G. B., Ramesh, M., Berg, T., & Learned-Miller, E. (2007, October). *Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments* (Technical Report No. 07-49). University of Massachusetts, Amherst. http://vis-www.cs.umass.edu/lfw/lfw.pdf

Marques, L. S., Gresse von Wangenheim, C., & Hauck, J. (2020). Teaching Machine Learning in School: A Systematic Mapping of the State of the Art. *Informatics in Education*, *19*(2).

Morsy, S., & Karypis, G. (2019). Sparse Neural Attentive Knowledge-based Models for Grade Prediction. *International Educational Data Mining Society*.

Nafea, I. T. (2018). Machine Learning in Educational Technology. In H. Farhadi (Ed.), *Machine Learning - Advanced Techniques and Emerging Applications*. IntechOpen. https://www.intechopen.com/books/machine-learning-advanced-techniques-and-emerging-applications/machine-learning-in-educational-technology

Nichols, P. T. (2012). Feedback in an Age of Efficiency. *Educational Leadership*, *70*(1), 71-74.

Park, S., & Ryu, J. (2019). Exploring Preservice Teachers' Emotional Experiences in an Immersive Virtual Teaching Simulation through Facial Expression Recognition. *International Journal of Human-Computer Interaction*, *35*(6).

Transparency Note Azure Cognitive Services: Face API [White paper]. (2019, May 6). Microsoft. Retrieved February 5, 2021, from https://azure.microsoft.com/mediahandler/files/resourcefiles/transparency-note-azure-cognitive-services-face-api/Face%20API%20Transparency%20Note%20(March%202019).pdf

Yang, X. (2019). Accelerated Move for AI Education in China. *ECNU Review of Education*, *2*(3).

Yu, K., Gong, R., Sun, L., & Jiang, C. (2019). The Application of Artificial Intelligence in Smart Library. *International Conference of Organizational Innovation (ICOI 2019)*, *100*, 708-713.

Zhang, H., Magooda, A., Litman, D., Correnti, R., Wang, E., Matsmura, L.C., Howe, E., & Quintana, R. (2019). eRevise: Using Natural Language Processing to Provide Formative Feedback on Text Evidence Usage in Student Writing. *Grantee Submission*.