

A Project Report On

Code-Kiet

Submitted for partial fulfillment of the
Requirements for the award of the degree of

Bachelor of
Technology

Computer Science and Engineering

Submitted by

Anshu Singh (1900290120018)
Ashish Kumar Sharma (1900290100039)
Ashutosh Dubey (1900290100040)

Under supervision of

Prof. Hriday Kumar Gupta



Dr. A.P.J. Abdul Kalam Technical University, Lucknow

DECLARATION

We hereby declare that this submission is our own work and that, to the best of our knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

Signature

Name:

Roll No.:

Date:

CERTIFICATE

This is to certify that Project Report entitled “Code-Kiet” which is submitted by Ashish Kumar Sharma, Ashutosh Dubey, and Anshu Singh in partial fulfillment of the requirement for the award of degree B. Tech. in Department of Computer Science & Engineering of Dr. A.P.J. Abdul Kalam Technical University, Lucknow is a record of the candidates own work carried out by them under my supervision. The matter embodied in this report is original and has not been submitted for the award of any other degree.

.

Date:

Supervisor Name:

(Designation)

ACKNOWLEDGEMENT

It gives us a great sense of pleasure to present the report of the B. Tech Project undertaken during B. Tech. Final Year. We owe special debt of gratitude to Prof Hriday Kumar Gupta, Department of Computer Science & Engineering, KIET, Ghaziabad, for his constant support and guidance throughout the course of our work. His sincerity, thoroughness and perseverance have been a constant source of inspiration for us. It is only his cognizant efforts that our endeavors have seen light of the day.

We also take the opportunity to acknowledge the contribution of Dr. Vineet Sharma, Head of the Department of Computer Science & Engineering, KIET, Ghaziabad, for his full support and assistance during the development of the project. We also do not like to miss the opportunity to acknowledge the contribution of all the faculty members of the department for their kind assistance and cooperation during the development of our project.

We also do not like to miss the opportunity to acknowledge the contribution of all faculty members, especially faculty/industry person/any person, of the department for their kind assistance and cooperation during the development of our project. Last but not the least, we acknowledge our friends for their contribution in the completion of the project.

Date:

Sign:

Name :

Roll No.:

ABSTRACT

In the modern world, everyone heard about coding, programming, and logic building. Those who are familiar with coding and different languages know the meaning of syntax in the programming languages. The spreading awareness of the use of programming languages and logic building made the criteria of different exams and different organization recruitment exams. The sense of the importance of this situation creates competition among the candidates who are preparing and giving online exams. The increasing competition also comes with loopholes in the system that the candidates found different types of cheating methods for the exams. They cheat in these kinds of online exams by different methods like copying the codes and not presenting in the online exam and in their place someone else gives the exams. In our project, we try to provide a solution to this problem by using real-time face recognition in these exams to examine the presence of the authenticated candidate. And examine their facial expression and recognize the give warning according to their activity. The other method is to use a good plagiarism checker to stop the code copying. In this research paper, we find and conclude the use of machine learning and plagiarism checking types and algorithm for this. In addition to this, we try to provide resources of different kinds of skills such as languages, subjects, etc. This research paper describes how facial recognition and plagiarism checking can be applied to online exams and assignments. This article first examines one of the most reliable new face recognition methods using various kinds of artificial neural networks (ANNs) and convolutional neural networks (CNNs). We then discuss how plagiarism is becoming a serious problem for the intellectual community. Plagiarism detection at various levels is a big topic. The problem is compounded when plagiarism is found in the source code. The source code may be in the same language or translated into another language. This kind of plagiarism can be found not only in academic papers, but also in industries dealing with software design. The main problem with source code plagiarism is that different programming languages may have different syntax. Consider different options for plagiarism checking of source code, regardless of programming language.

TABLE OF CONTENTS	Page No.
DECLARATION.....	ii
CERTIFICATE.....	iii
ACKNOWLEDGEMENTS.....	iv
ABSTRACT.....	v
LIST OFFIGURES.....	ix
LIST OF TABLES.....	xi
LIST OF ABBREVIATIONS.....	xii
 CHAPTER 1(INTRODUCTION).....	 1
1.1.Introduction.....	1
1.2. Project Description.....	1
 CHAPTER 2 (LITERATURE RIVIEW).....	 2
2.1. Sub heading1	2
2.2. Sub heading2	2
2.2.1 Sub sub-heading1	2
 CHAPTER 3 (PROPOSED METHODOLOGY)	 3
3.1. Sub heading	3
 CHAPTER 4 (RESULTS AND DISCUSSION)	 3

CHAPTER 7 (CONCLUSIONS AND FUTURE SCOPE).....	5
7.1. Conclusion.....	5
7.2. Future Scope.....	5
REFERENCES.....	6
APPENDEX1.....	7

LIST OF FIGURES

Table . No.	Description	Page No.
1	Block Diagram of Face detection	
2	ANN	
3	Feed forward neural network	
4	Types of Plagiarism	
5	Types of Textual Plagiarism	

CHAPTER 1

1.1 INTRODUCTION

In our project we try to develop a website that provide the following features

1. Real time face detection for the online contest
2. Plagiarism checking implementation to reduce the cheating and copying of the content
3. Users updated performance and profile
4. Resources like
 - Problems
 - Subjects references
 - Skill development paths etc

1.2 PROJECT DESCRIPTION

1. Use of the different type of the artificial neural network for the analyzing of the input in the form of images, patterns etc. For this purpose we found that the Convolutional neural networks (CNNs). CNN are used for mostly recognizing the images, patterns, and data coming in the form of grids like the pixels informations. And it has different layers namely[INPUT-HIDDEN LAYER(CONV-RELU-POOL-FC)-OUTPUT].They work in hand in hand for the output.
2. After that We then discuss how plagiarism is becoming a serious problem for the intellectual community. Plagiarism detection at various levels is a big topic. The problem is compounded when plagiarism is found in the source code. The source code may be in the same language or translated into another language. This kind of plagiarism can be found not only in academic papers, but also in industries dealing with software design. The main problem with source code plagiarism is that different programming languages may have different syntax. Consider different options for plagiarism checking of source code, regardless of programming language.
3. We have gathered that information from the student who are expert in there respective domain that there are a lot of learning resources for students available across different platforms which has created a sense of ambiguity. Nowadays, due to the availability of large amount of resources in an unorganized manner the students face difficulties while preparing for competitive exams, interview preparations or building a project.
4. So we try to build a website that can organize some contest with the face detection and plagiarism and resource hub

An approach of using face detection mechanism and plagiarism checking algorithms in proctored exams

Prof. Hriday Kumar Gupta, Ashish Kumar Sharma, Ashutosh Dubey,
Anshu Singh

KIET Group Of Institutions, Delhi NCR, Ghaziabad – 201206

Part 1.1: Introduction

Face detection and real-time facial recognition play an important role in applications such as robotic intelligence, smart cameras, security monitoring or even criminal identification [11]. Conventional face detection and face recognition algorithms are designed for static face images or color images. Using facial recognition method, we can identify or verify the identity of an individual using their face. Facial recognition has been used for various applications such as automated classroom attendance management systems [15]. Applying real-time facial recognition systems to online websites for various purposes in a very accurate and large scale is always a challenge [13]. There are several approaches to deal with recognition problems. Since faces are usually round or oval with the same color, one of the simplest approaches is to use color segmentation to detect faces. However, the

use of color segmentation cannot adapt to changing environments such as lighting conditions. More adaptive and robust methods may not be able to work in real time because they require more computing power [11].

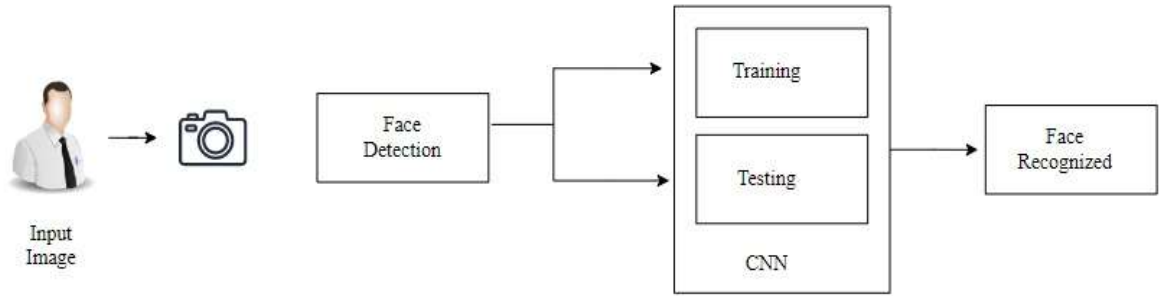


Fig 1: Block diagram of face detection mechanism.

1.2 About artificial neural networks.

Neural networks are also called as artificial neural networks (ANNs) or simulated neural networks (SNNs), are a subset of machine learning and are the content of deep learning algorithms. Their name and structural design are inspired by the human brain and try to mimic the way biological neurons in the brain signal to each other [14].

Artificial neural networks (ANNs) consist of layers of nodes that contain an input layer, one or more hidden layers, and an output layer or target layer. Each node or artificial neuron connects to one another and has an associated weight(w_i) and threshold. If the output of any individual node is above a particular specified threshold, that node is

activated and sends data to the next layer of the network. Otherwise, nothing will be passed to the next network layer [14].

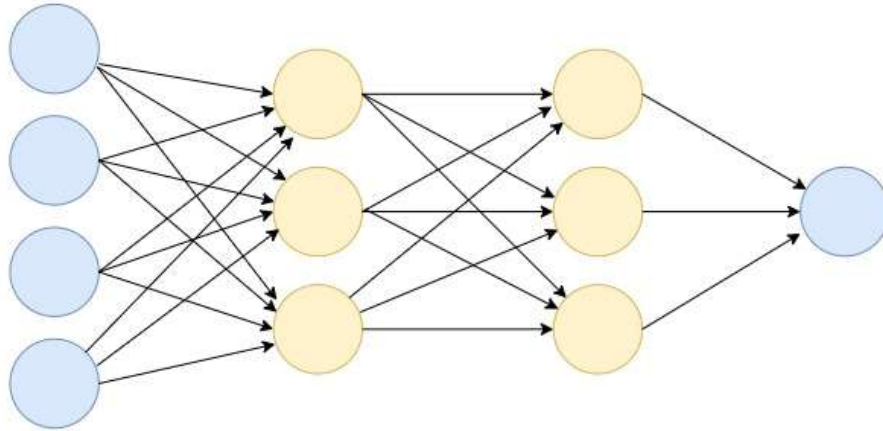


Fig 2 : Artificial neural network

1.3 Feed forward neural network

Feed-forward neural networks allow signals to travel in one direction only, from input to middle layer to output. There is no feedback such as the output of some layer does not influence that same layer. Feed-forward networks are the simplest networks that associate inputs with outputs. It can be used in pattern recognition, face recognition, and in other things. Each unit in the hidden layer is generally completely connected to some units in previous the input layer. The middle units in the hidden layer compute their output by multiplying the value of each input by its correlating weight(w_i), inserting these values, and using the transfer function. A neural network can have several hidden layers, but one hidden layer is adequate. The wider the layer the higher the capacity of the network to identify designs, faces, and patterns. The final unit on the right is the output layer (target layer) because it is

linked to the output of the neural network [14]. It is completely connected to units in the hidden layer. The neural network is generally used to compute a single value (target value), therefore there is only one unit in the output layer and the value [9].

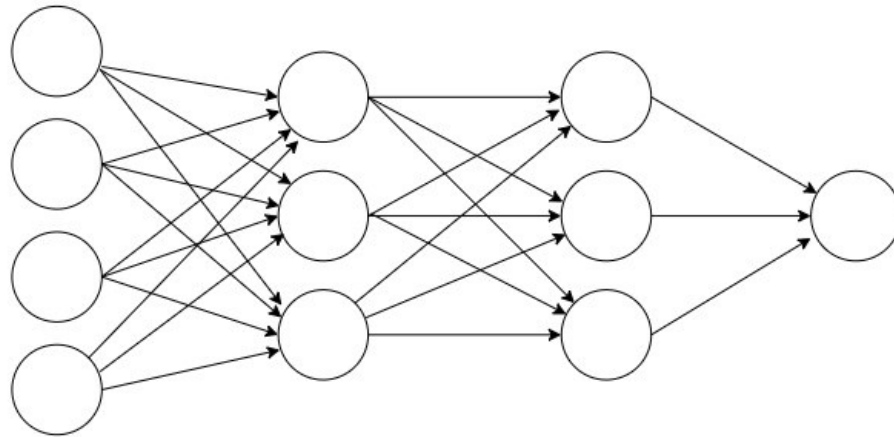


Fig 3: Feed- forward neural network

1.4 Convolutional Neural Networks (CNNs) architecture

Convolutional Neural Networks (CNNs) are similar to other neural networks but provided that inputs are images that allow designers to encode certain properties of that feature into the architecture. The CNN architecture includes a sequence of layers with the simplest architecture [INPUT-HIDDEN LAYER(CONV-RELU-POOL-FC)-OUTPUT]. INPUT layer contains the raw pixel values of the images, the CONV layer consists of a fixed-size filter that shifts window pattern to perform a convolution operation on the windowed image to extract the desired features [9]. Padding is used on the size of the input images to overcome irregular mapping with filter size. RELU stands for rectified linear units, which is an element-by-element activation function that assigns the hidden units a value of zero, POOL indicates pooling layer that is responsible for downsampling and dimensionality reduction,

which in turn reduces computational energy that is required to process data. The pooling layer also has a feature that scrolls like an input window extract important features that are rotationally and positionally invariant. Maximum pooling and average pooling are two common functions used in it. FC is a fully connected layer in which every input neuron is connected to every other neuron at the output, and this layer is responsible for computing the score of the given class, resulting in N outputs where N indicates the number of classes or categories to be classified. The class with the highest number of points is designated as the class an assumed CNN architecture class [9]. The FC layer is also referred to as the DENSE layer. It may be noted that CNN the architecture can be modified based on system design and performance requirements. And others the layers that are used in the CNN architecture include DROPOUT and FLATTEN. The DROPOUT layer is a regularization a technique to avoid the CNN overfitting problem where a fraction of the inputs (also referred to as the dropout rate) are dropped by setting their values to 0 at each update during training. The values of the inputs that are preserved are scaled so that their sum remains the same during training. FLATTEN layers are introduced just before the FC layer to convert two-dimensional features into one dimension for easy computation [9].

2.1 Introduction

Just as the Industrial Revolution changed the way we live and work, the development of computers, the Internet, and web search engines changed the way we think. Vast amounts of information will be available to anyone with internet access, and relatively few skills will be required. Therefore, traditional methods of searching for information from books may become a thing of the past. The use of the Internet and online resources also poses some serious problems, as there is no way to prevent users from using the Internet. In fact, the Internet provides tools that facilitate plagiarism, and plagiarism plagues

many fields. Education is no stranger to this. Traditional education is now complemented by online resources, web classrooms, virtual worlds, and easy online access to source documents. Many researchers report that plagiarism is on the rise and a serious problem in education. Educational institutions are now fighting for more (better) students, but not increasing their workforce. In addition, the upcoming Bologna process will reduce contact times. As a result, teachers are less able to identify students who need additional help, and assessment scores are the only indicator of whether someone needs help. Therefore, increasing the number of students reduces (further) the energy spent on a single assessment. Unfortunately, students are aware of this and there are some cheats (easier) for that. If someone else's assessment is copied and the teacher doesn't find it, the student may leave the course with insufficient knowledge, and this may cause problems later in life. More worryingly, research shows that not only do students regularly plagiarize, but they also believe it's okay to plagiarize. Plagiarism in a copyrighted work includes the use of language (communications), opinions, results, texts, graphic designs, computer-related programs, drawings, diagrams, graphics, works of art, knowledge, teachings on paper, electronic works, or other works created. Created and presented by others consisting of contemporary and fresh works. Plagiarism can be described as computer spam, hacking, computer viruses, phishing, piracy, and other criminal-like electronic errors. Plagiarism can be described as copying the work of another person in whole or in part without citation. Plagiarism involves copying the original data from a journal, research paper, etc. and pasting it directly into your code with little or no modification. When students and researchers submit projects and work, it is expected that most of them will copy the project or work from other sources, as almost all data is available on the Internet these days. It's easy to use a search engine to get data about the topic names you use in your work without naming the author. This makes it essential to detect or avoid plagiarism for educational purposes. Plagiarism is a difficult task in some areas of education. Therefore, in all educational fields, removing plagiarism requires the use of tools that can detect plagiarism. The rest of the article is

organized as follows. Describes how to detect plagiarism. Discusses various applications of plagiarism detection. We will discuss the different approaches available. A comparative analysis of various research papers is presented. Discuss problems and challenges. Like all similar faculties, due to the spread of ICT [17], the Faculty of Education at the University of Ljubljana is also facing this problem. As a result, we started a project to study plagiarism. These include why and how students plagiarize, researching known plagiarism detection systems, and developing a framework for professors and assistants to detect plagiarism between different file types. By understanding the reasons behind plagiarism, we aim to develop an electronic plagiarism detection service that is more useful than general systems. A secondary goal is to prepare plagiarism advice and materials for teachers and students that disseminate the following ideas. The structure of the paper is as follows. It discusses the problem of plagiarism, how plagiarism stands out in programming courses, why some students plagiarize, and detection systems that can be used to detect plagiarism [17].

2.2 Plagiarism and its type

The term "plagiarism" has many definitions. The Encyclopædia Britannica [19] defines plagiarism as "the act of passing on the work of others as your own. This fraud is closely related to counterfeiting and piracy, which commonly violate copyright law." Similarly, the web Starr's dictionary defines plagiarism as "text copied from another and presented as one's own work." plagiarism detection methods plagiarism detection methods have his two types: text plagiarism and source code plagiarism, as shown in Figure 1

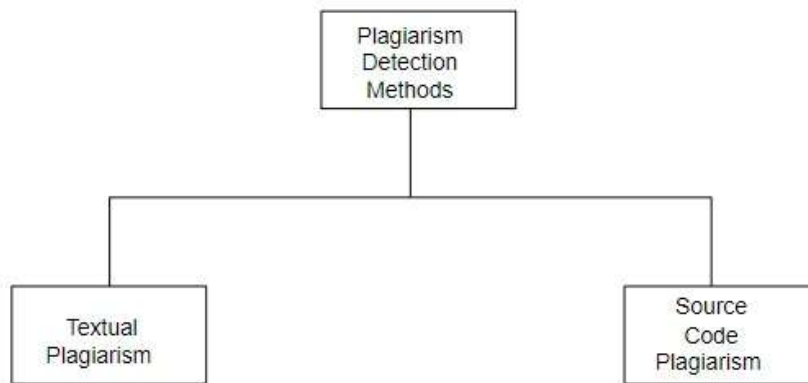


Fig 5 . Types of Plagiarism detection methods

Types of Plagiarism Detection Methods

A. Text plagiarism Text plagiarism. This type of plagiarism is mostly done when students write term papers and when researchers write research papers. To detect this kind of plagiarism, researchers have developed many tools that can detect plagiarism automatically. Different types of text plagiarism are shown in Figure 1. 2 and described below [4].

1) Grammar-based method: This method is used to identify plagiarism based on the grammatical structure of the text. Works for direct copy and paste, but not for detecting modified text.

2) Semantics-based methods: based on the vector space model. In this method, the author proposes counting the number of plagiarized words in the document. It works well with non-partial documents because it is difficult to correct the position of copied text within a complete document.

3) Grammar-semantics Hybrid method: This method is a valid approach that can recognize synonyms or paraphrases of words. In this way, we overcome the shortcomings_{1c} of both grammar-based and semantics-based methods.

4) External Plagiarism Detection Method: This method searches for plagiarism in text based on a data set. A record is a collection of documents in which the contents of sentences or parts of sentences are copied.

5) Clustering-Based Plagiarism Detection Methods: Document clustering is used for information retrieval for a variety of reasons. Document clustering can be used for plagiarism detection to reduce search time.

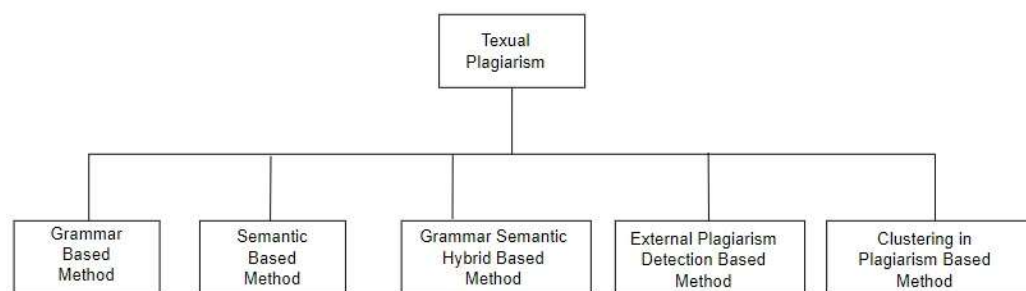


Fig 6. Types of textual plagiarism.

2.3 Reasons for plagiarism

There are many reasons why students may use each other's work or collude in doing a particular task [20].

- Working closely with a colleague, a struggling student writes a paper that they believe will be accepted.
- A timid student secretly copies and edits a colleague's program with or without permission.
- Unmotivated (but not necessarily weak) students copy their peers' programs and edit them with the intention of minimizing the work required. At the individual student level, he often explains why certain

individuals commit material plagiarism in three categories. The three categories [17] relate to the student's personal situation, personal characteristics, and the availability of means and opportunities for plagiarism. Students can find and download materials from a variety of sources with little reading, effort, or creativity required. In addition, web services containing such material are customized, making them difficult for anti-plagiarism web crawling software to detect [18]. The Internet undoubtedly encourages plagiarism, but it lacks the moral force to incite cheating to otherwise honest students. Lack of rules and prosecution in cases of plagiarism can encourage students to indulge in practice is known. However, the most powerful motivation for student misbehavior (according to Bjorklund and Wenestam [18]) is the desire to get good grades, which can depend on other considerations. For example, due to an innate need to prove one's worth to oneself and the world, or a pathological fear of failure. Individual Circumstances: For example, students who need to work to cover their tuition fees may have less time to study and may need to compress their available study time due to the heavy academic burden. Due to time pressure, more and more students may resort to plagiarism. It is also interesting that men are more likely to admit to plagiarism than women. The standard "stupid" attempt to cheat program assignments is to get a copy of a working program and change statement spacing, variable names, prompts, and comments. Comparing every pair of solutions for evidence of plagiarism seems to be the approach to uncovering this kind of fraud. But even the above case is almost enough to require careful manual comparison, which simply becomes impractical for large classes [17]. Because it's usually more than just a few tasks, programming classes urgently need automated tools that perform reliable and objective plagiarism detection. In programming courses, he has two sources of solved tasks. Internet and other students. For the web, products like Turnitin and PlagiServe use the web to detect plagiarism. This is what search engines like Google have been doing for a long time. However, the second source, other students, is more common and requires a separate plagiarism detection system. Attempting to assess plagiarism using technical means makes it

difficult to recognize differences between different types of text. Much work has been done in the past to discover correspondences between texts. A plagiarism detection method should provide a measure that quantifies how close two program source codes are. Except for literal copies, recognition approaches that use direct comparison of text files are clearly fragile. Of course, a simple "diff" file will only detect the most obvious cheating attempts. There are various electronic systems for detecting plagiarism in programming courses. From the mid-1970s to the late 1980s, systems for finding similarities by counting and comparing program attributes became popular. This technique was called attribute counting. Later, plagiarism detection systems were introduced that examined and compared program structures [5]. A study of standard software metrics and redundant code was also used. There are even servers on the web that detect plagiarism. For example, his JPlag at the University of Karlsruhe tries to find pairs of similar programs, and his MOSS server at Berkeley looks for similar code sequences in many programs [17]. Each system creates a web page where instructors can see which ones are suspiciously similar. All of these techniques work by running an analyzer on a group of submissions to detect similarities and calculate the likelihood of plagiarism. Many approaches take a lexical approach and classify program tokens as language keywords and user symbols [5]. A simple plagiarism detection system converts a source program into tokenized strings and uses dynamic programming to compare the strings. While successful in pointing out couples or groups that have submitted similar work, it is limited in identifying the original author of the work

RESULT AND FUTURE SCOPE

1. After successful development and testing of the app we can provide
 - Local contest organizing web portal for college students
 - Help to evaluate the growth and progress of the student
 - Help to manage the student detail
 - If possible, we can provide the Alumni connection feature
 - Helps the student to practice the contest environment
 - Reduces the time for the student to target the weak points
 - Helps the students to learn the skills in the right manner without any kind of ambiguity in the resources.

REFERENCES

1. Target, S. (1995) Poster campaign warning against plagiarism. Times Higher Education Supplement, No. 1159 pi
2. (Anonymous). University Combats Plagiarism, Network, Ausgabe 19, S. 6, University of Wolverhampton, 1995
3. (Anonymous). Plagiat or Pastiche. Times Higher Education Supplement, Nr. 1159 Pi 1
4. Leech, M. (1995) A Critical Evaluation of Existing Literature Concerning Plagiarism Detection Systems with FoxPro, unveröffentliche MSc - Dissertation, University of Wolverhampton
5. Parker, P. and Hamblen, J. O. (1989) Computer Algorithms for Detecting Plagiarism. IEEE Transactions on Education, Vol. 32, No. 2, pp. 94-99 .
6. Salt, N.F. (1982) Defining Software Science Counting Strategies SIGPLAN Notices, 17, p58-67
7. Berghel, H. L., Sallachi, D. L., (1984) Measurements of Program Similarity in Identical Task Environments, SIGPLAN 19(8) p65-76
8. Oman, P. W, Cook, C. R, (1989) Author Analysis of Programming Styles. ACM pp320
9. Pranav KB and Manikandan j Design and Evaluation of a Real-Time Face Recognition System using Convolutional Neural

Crucible of Research and Innovation (CORI) and Department of ECE, PES University.

10. Seema Lalwani, Hiteshi Vanjani and Mona Mulchandani Understanding Real-Time Face Detection Techniques Jhulelal Institute of Technology.
11. Xin Zhang, Thomas Gonnot, Jafar Saniie Real-Time Face Detection and Recognition in Complex Background, Illinois Institute of Technology.
12. Arief Agus Sukmandhani Computer Science Department, Binus Online Learning, Arief Agus Sukmandhani Computer Science Department, Binus Online Learning Bina Nusantara University Jakarta.
13. Zhu, X. and Ramanan, D. (2012) Face Detection, Pose Estimation and Landmark Localization in the Wild. IEEE Conference on Computer Vision and Pattern Recognition, Providence, 16-21 June 2012, 2879-2886.
14. IBM Cloud ANN - <https://www.ibm.com/en/cloud/learn/neural-networks>
15. Fahad P, Md. Mahmudul, Md. Atiqur, Susan M, Moslehuddin M, and Pandian V. (2017) “Face recognition based real time system for surveillance.” Intelligent Decision Technologies, IOS Press, 11 (2017): 79–92.
16. Cha Zhang and Zhengyou Zhang , A Survey of Recent Advances in Face Detection.
17. Dejan Sraka, Branko Kaučič Faculty of Education, Kardeljeva pl. 16, 1000 Ljubljana.
18. Bennett R. Factors associated with student plagiarism in a post-1992 university. Assessment & Evaluation in Higher Education 2005, 30(2): 137-162
19. plagiarism - Britannica Online Encyclopedia, <http://www.britannica.com/EBchecked/topic/462640/plagiarism> [02/16/2009].
20. Schiller R M. E-Cheating: Electronic Plagiarism. Journal of the American Dietetic Association 2005; 105 (7): 1058-1062.