SOFTWARE PLAGIARISM DETECTION IN MULTITHREADING USING MACHINE LEARNING

Presented by : Anju Sreekumar
S6 MCA
Roll No:08
TVE18MCA009

Guided by : Dr.Sabitha S

June 28, 2021

COLLEGE OF ENGINEERING TRIVANDRUM

OVERVIEW

INTRODUCTION

- The Software plagiarism, which arises the problem of software piracy is a growing major concern nowadays.
- The customers may develop a modified version of the original software in other types of programming languages.

LITERATURE REVIEW

- Semantics-preserving code obfuscations.
  Developed by Lannan Luo,combines rigorous program semantics with longest common subsequence based fuzzy matching.
- Software watermarking.
  Collberg and Thoborson proposed software watermark technique By embedding a unique identifier, i.e., a watermark.
- Software birthmarking.
  Is illustrated by Myles and Collberg,birthmarking attempts to extract a set of characteristics that can be used to uniquely identify a program.

MOTIVATION

- The Software plagiarism,It is a serious risk to the software industry that gives huge economic damages every year.
- GPL (GNU General Public License) allows users to modify GPL compliance programs freely, as long as the derivative works also follow the tenets of GPL
- However, driven by commercial interests, some companies and individuals incorporate third party software without respecting the licensing terms.

PROPOSED SYSTEM

- Proposed a methodology for software plagiarism detection in multi-programming languages based on machine learning approaches.
- Software birthmarks have been proposed as a method for enabling the detection of programs that may have been stolen by measuring the similarity between the two software.

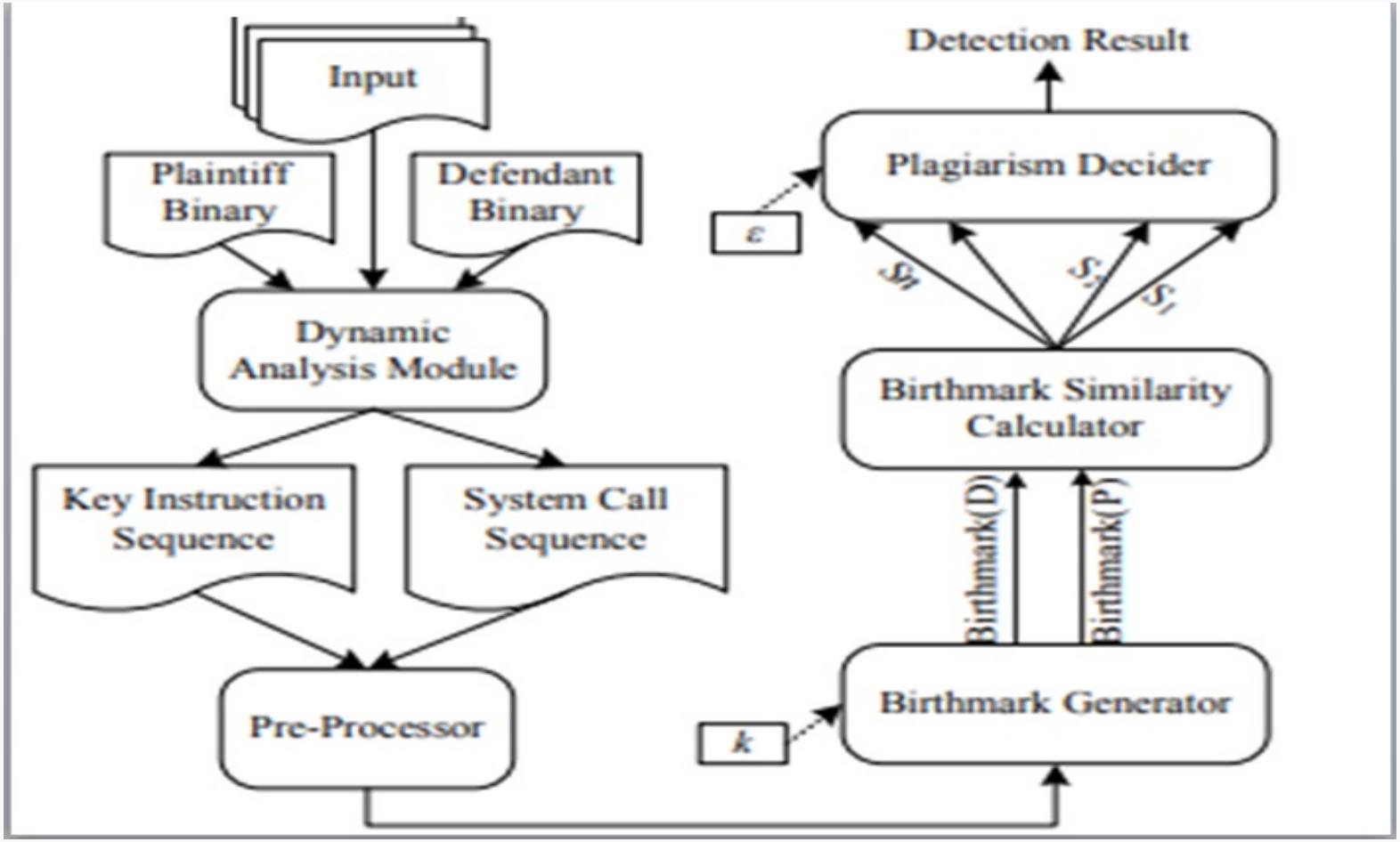
Experimental Evaluation

- Hardware Requirements
  Processor : Intel Core i3
  Memory : 4 GB RAM
- Software Requirements
  Operating System : Windows
  IDE: Visual Studio

MODULES

- Birthmark creation
- Similarity Calculation
- Plagiarism detection

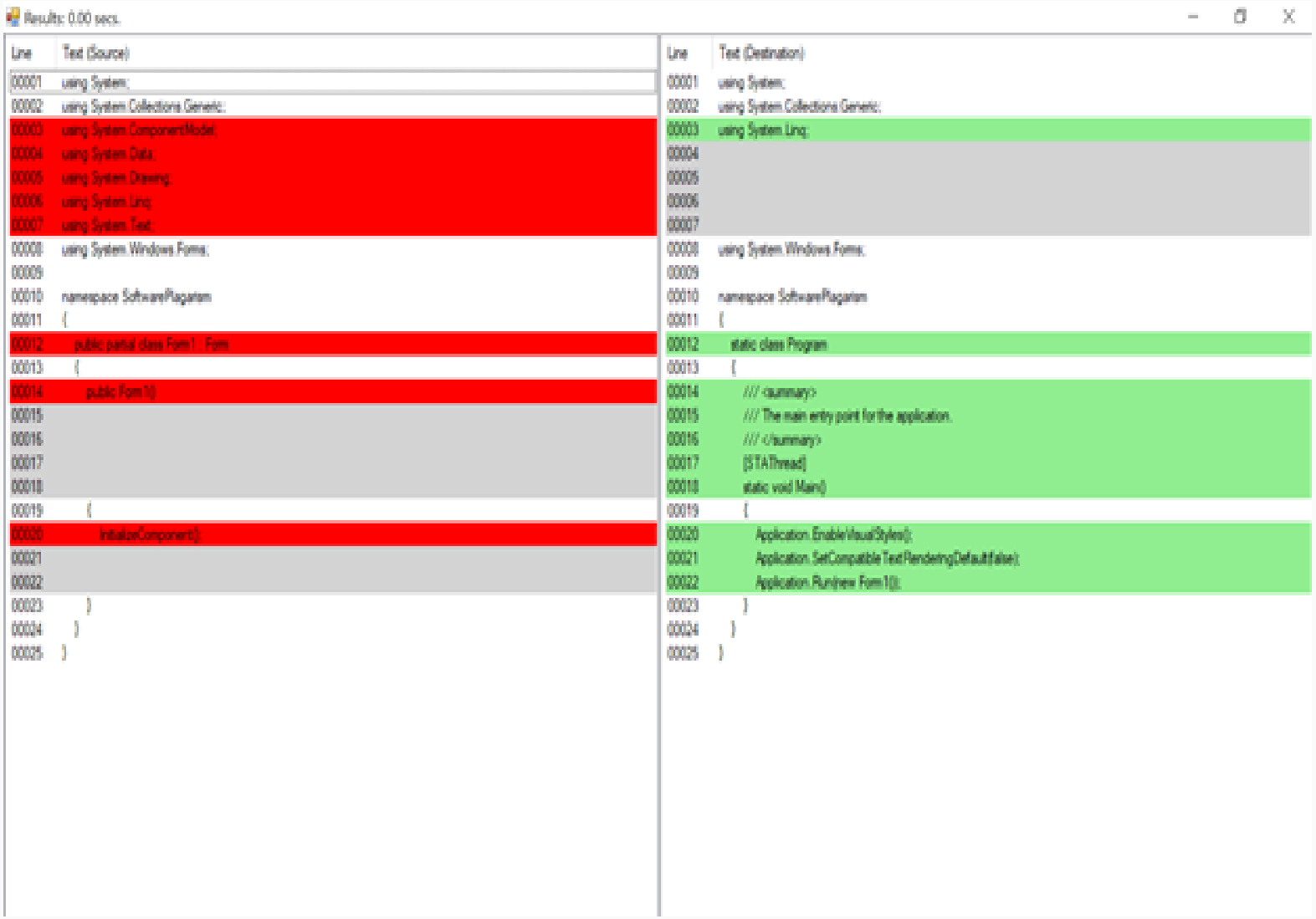
MODEL


Input
Plaintiff Binary
Defendant Binary
Dynamic Analysis Module
Key Instruction Sequence
System Call Sequence
Pre-Processor
Detection Result
Plagiarism Decider
ε
Birthmark Similarity Calculator
Birthmark(D)
Birthmark(P)
Birthmark Generator
k

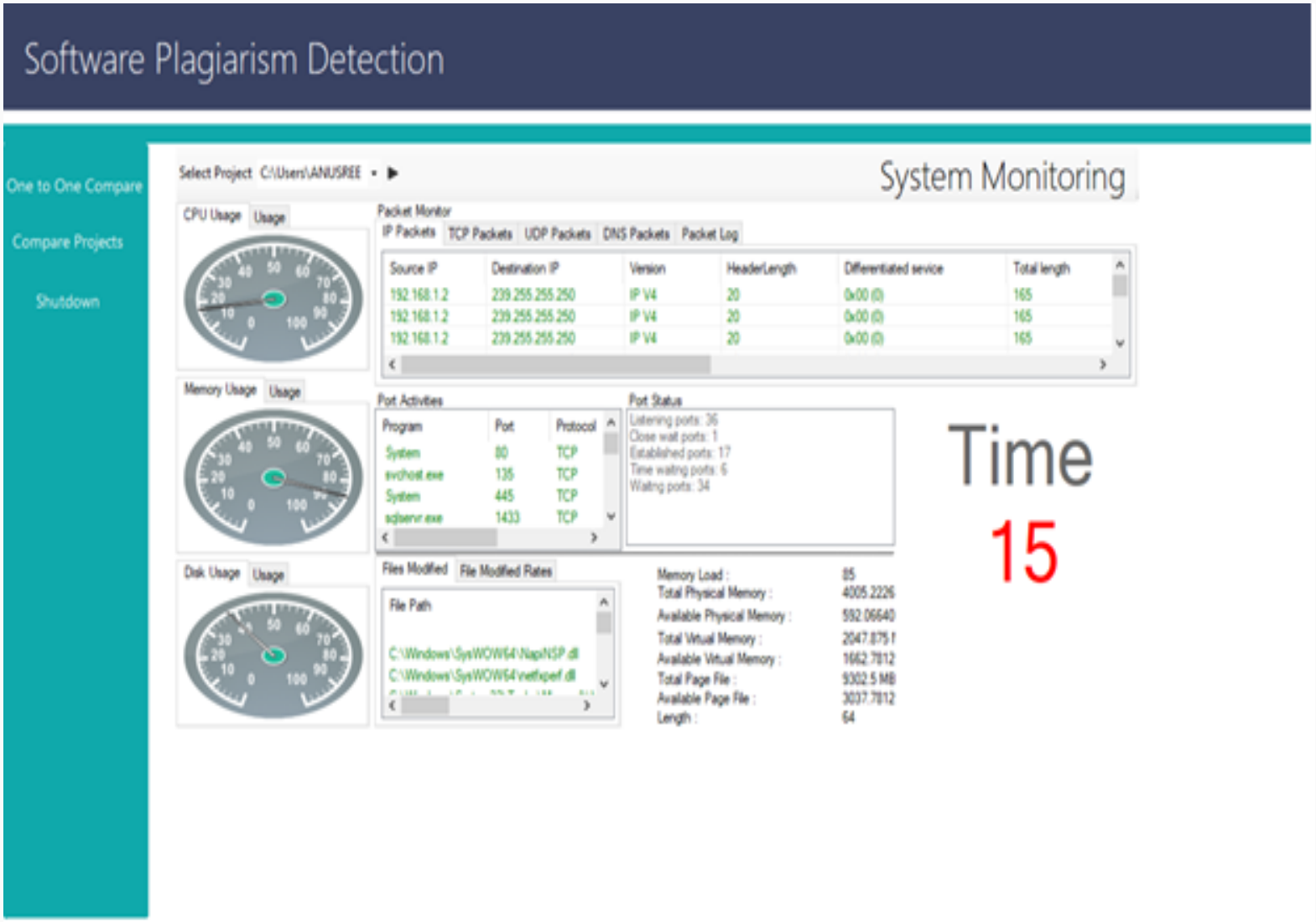
SCREEN SHOTS

Login page

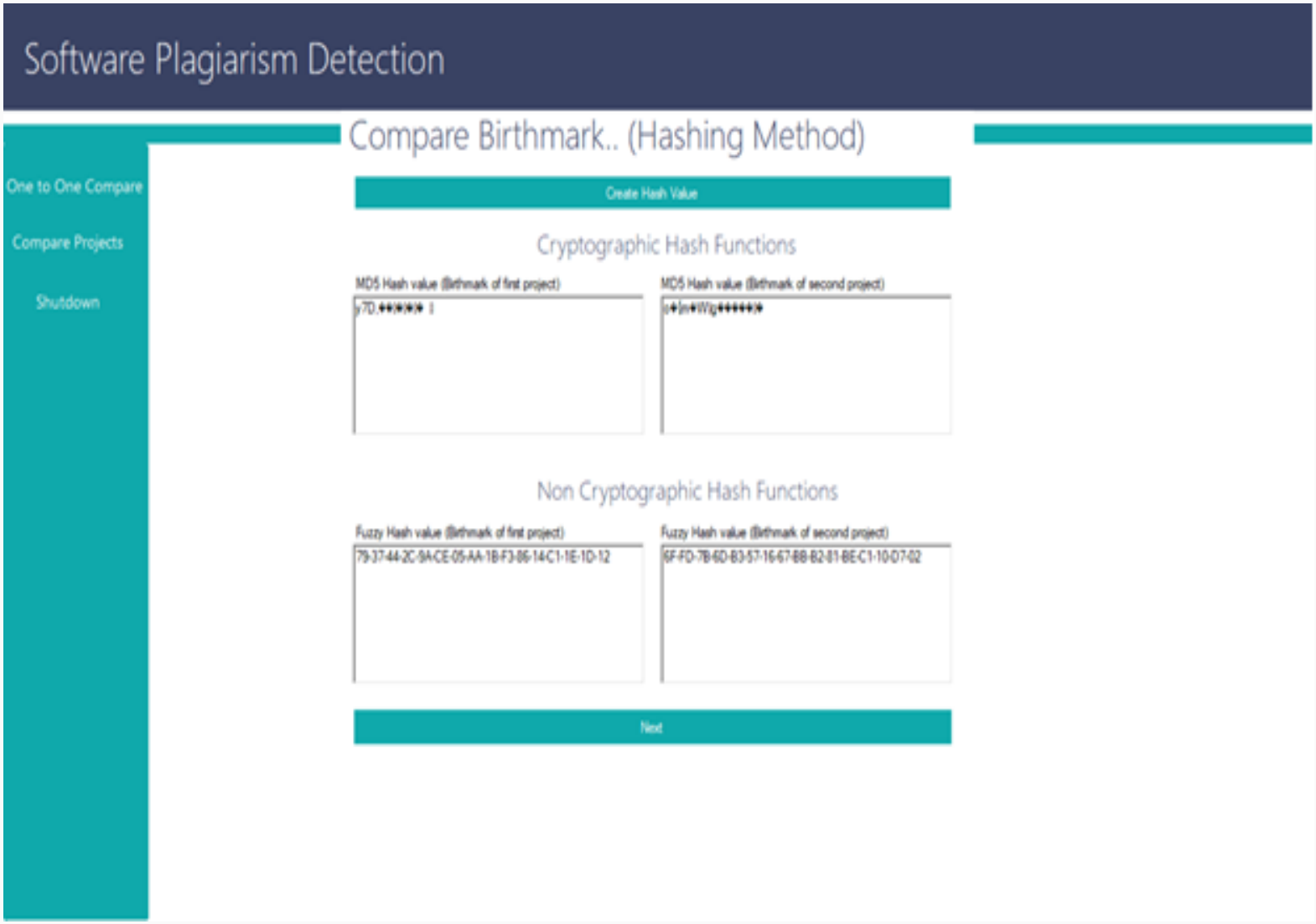
Result of file comparison



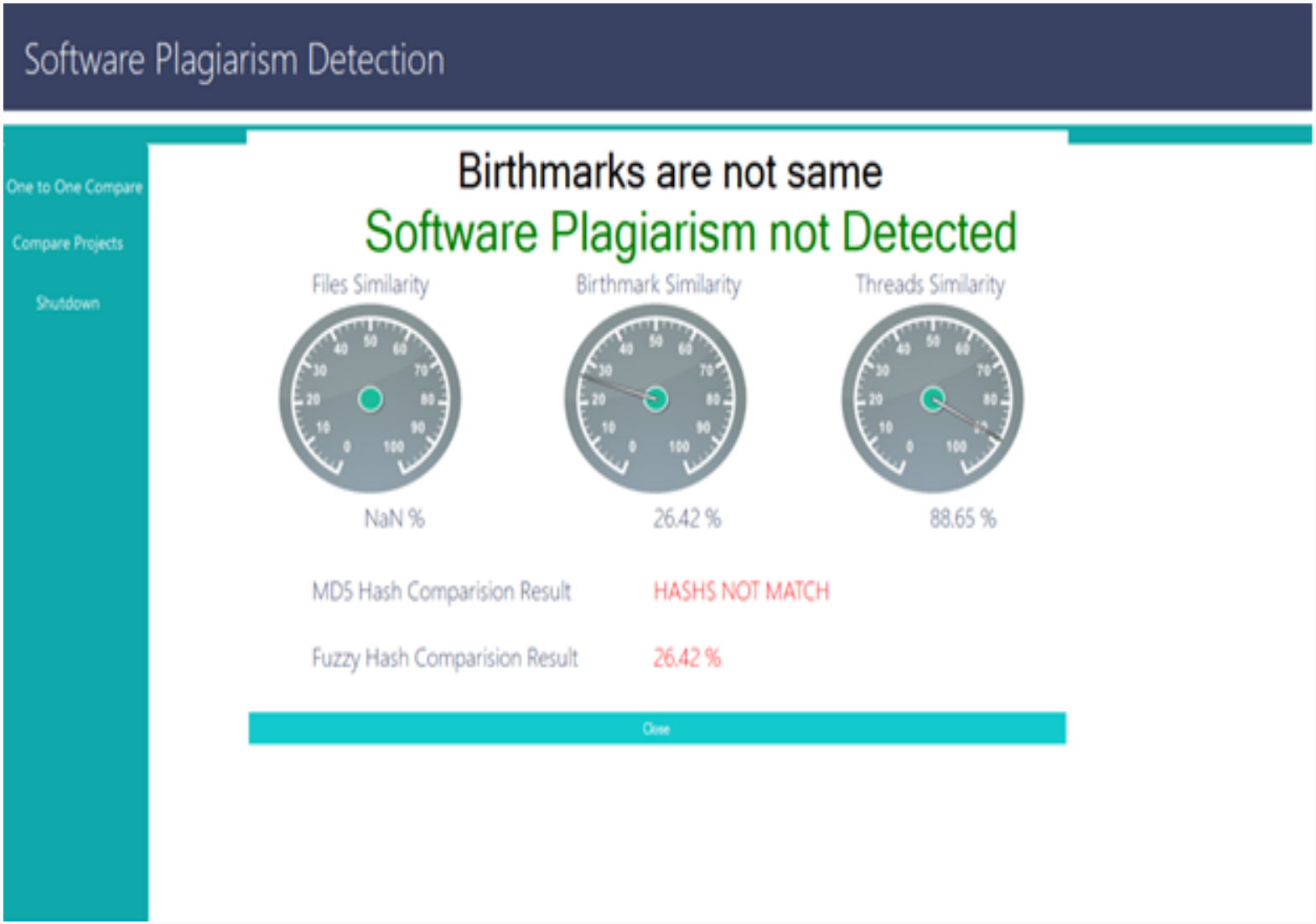
System monitoring for birthmark generation



Applying hash functions



Plagiarism detection



Result and Discussion

- Each of the features, functionalities are working correctly. This proposed system have met all the proposed requirements in the proposed system.
- Expected software plagiarism detected perfectly

Future Scope

- It can extend my work in the next higher level by enhancing our model to generate software plagiarism, comparing with real world plagiarism cases.
- Identifying pairs of applications that plagiarism may exist is extremely laborious. So this can be taken as one of the future work.

Conclusion

- The proposed system is for software plagiarism detection in multi-programming languages based on machine learning approaches.
- This is completely a Machine Learning project, which makes use of fuzzy hashing technique.

REFERENCE

 Xi Xu,Ming Fan "Revisiting the Challenges and Opportunities in Software Plagiarism Detection", IEEE May 31,2020

THANK YOU