**NATIONALITY PREDICTION BY USING MACHINE LEARNING**

BY

**Nazmus Sakib**

**ID:162-15-8104**

**Md. Anowar Hossain**

**ID:162-15-7762**

**Md. Ismail Hossain Sobuj**

**ID:162-15-7910**

This Report Presented in Partial Fulfilment of the Requirements for the degree of Bachelor of Science in Computer Science and Engineering

Supervised By

**Sheikh Abujar**

Lecturer (Senior Scale)

Department of CSE

Daffodil International University

Co-Supervisor By

**Md. Tarek Habib**

Assistant Professor

Department of CSE

Daffodil International University

**DAFFODIL INTERNATIONAL UNIVERSITY**

**DHAKA, BANGLADESH**

**DECEMBER 2020**

**APPROVAL**

This Project titled “Nationality Prediction by using machine learning” submitted by Nazmus Sakib, Md. Anowar Hossain & Md. Ismail Hossain Sobuj to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on **Date Must**

**BOARD OF EXAMINER**

**Name Chairman**

**Designation**

Department of CSE

Faculty of Science &Information Technology

Daffodil International University

**Name Internal Examiner**

**Designation**

Department of CSE

Faculty of Science & Information Technology

Daffodil International University

**Name External Examiner**

**Designation**

Department of ------

Jahangirnagar University

**DECLARATION**

We hereby declare that, this project has been done by us under the supervision of Name, Designation, Department of CSE Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma

**Supervised by:**

**Name**

Designation

Department of CSE

Daffodil International University

**Co-Supervised by:**

**Name**

Designation

Department of CSE

Daffodil International University

Submitted by:

Nazmus Sakib

ID:162-15-8104

Department of CSE

Daffodil International University

Md. Anowar Hossain

ID:162-15-7762

Department of CSE

Daffodil International University

Md. Ismail Hossain Sobuj

ID:162-15-7910

Department of CSE

Daffodil International University

**ACKNOWLEDGEMENT**

First, we our heartiest thanks and gratefulness to almighty God for His divine blessing marks us possible to complete the final year project/internship successfully.

We rally grateful and wish our profound our indebtedness to **Sheikh Abujar**, Lecturer(Senior Scale), Department of CSE Daffodil International University, Dhaka. Deep knowledge & keen interest of our supervisor in the field of machine learning, natural language processing and data mining to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to Dr. Syed Akhter Hossain, Head, Department of CSE, for his kind help to finish our project and also to other faculty members and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while complete the course work.

Finally, we must acknowledge with due request the constant support and patients of our parents.

**ABSTRACT**

Machine learning is a branch of artificial intelligence that employs a variety of statistical, probabilistic and optimization techniques that allows computers to “learn” from past examples and to detect hard-to-discern patterns from large, noisy or complex data sets. Our work attempts to solve the problem of nationality prediction of humans based on their facial features. We have data from more than 60 countries all over the world. The performance of a system depends mainly on the feature extraction step because characterizing features makes it possible to distinguish between country to country. Our program analysis that a person is male or female and height, weight and facial features as ethnicity, age, hair color and eye color. By analyzing of all most 84 thousand of data sets, it can predict a person’s nationality. The accuracy of our program obtained using a MLP approach was 80% whereas the accuracy obtained by using a convolution neural network was a significant 98%. All   summary   similarity data measurement cases using the method provided an effective value and optimal result.

Keyword- data-summarization, Machine Learning, data Mining, Data Pre-Processing, Data Result Prediction

**TABLE OF CONTENTES**

**CONTENTS PAGE**

Board of examiners

Declaration

Acknowledgements

Abstract

**CHAPTER**

**CHAPTER 1: INTRODUCTION**

* 1. Introduction
  2. Motivation
  3. Rational of the study
  4. Research Question
  5. Expected Outcome
  6. Report Layout

**CHAPTER 2: BACKGROUND**

2.1 Related Work

2.2 Comparative Analysis and Summary

2.3 Scop of the Problem

2.4 Challenges

**CHAPTER 3: RESEARCH METHODOLOGY**

3.1 Research Subject and Instrumentation

3.2 Dataset Utilized

3.3 Statistical Analysis

3.4 Proposed Methodology

3.5 Implementation Requirements

**CHAPTER 4: EXPERIMENTAL RESULT AND DISCUSSION**

4.1 Experimental Setup

4.2 Experimental Result & Analysis

**CHAPTER 5: SUMMARY, CONCLUTION, RECOMEDATION AND IMPICATION FOR FUTURE RESEARCH**

5.1 Summary of the Study

5.2 Conclusions

5.3 Implication for Further Study

**References**

**LIST OF FIGURES**

**LIST OF TABLES**

**CHAPTER 1**

**Introduction**

* 1. **Introduction**

For the past few decades, facial analysis is one of the most research section in the field of computer vision and pattern recognition. Facial features of a person provides a variety of information like gender, age, ethnicity, hair color, eye color. In all over the world there are total 195 countries that at complicates to recognize nationality of a person. Because of some countries have almost same facial features, some have identically difference and also some have mixed facial features. So, height and weight of a person could be some helpful to make difference from country to country. As the people belonging to same country will have almost similar characteristics.  Similarly, people belonging to different country will have distinguishing feature to identify.

Histogram of Oriented Gradients (HOG) is the part of Machine Learning (ML) descriptor for object recognition is nearly a decades old, it is still intensely utilized today and with fantastic results. Histogram of Oriented Gradients for human detection demonstrated that the Histogram of Oriented Gradients (HOG) image descriptor and a linear Support Vector Machine (SVM) is highly preferred by many as it produces significant accuracy with less computation power. SVM can be used for linear regression and logistic regression classification tasks could be used to train highly accurate object classifiers.

In our program, there are 30 countries have been considered. To determine a person’s nationality, we have taken a person’s gender, height, weight, age, ethnicity, hair color, eye color. Our model can determine which country is he or she from by analyzing their characteristics with almost accurate percentage.

**1.2 Motivation**

Researchers who have previously researched this site on people from certain countries and nations have collected datasets of facial point of people in that country, such as the length of the nose, the width of the eyebrows, the color of the eyebrows, the color of the hair, and so on.

Where he predicts ethnicity is only the people of Asia and got a good result but the main problem here is to do face detection of married people because some points is the face of unmarried people would change after marriage which would reduce the accuracy of face detection.

The other part was done on European people. According to the previous work most of the European people have 90 to 95 percent similarity in face so that the length of the nose is almost equal to the width of the nose. Even the color of the eyebrows is the same as the color of the eyebrows.

Then we can easily predict nationality and if only Asian and if we use Histogram of Orientation Gradients (HOG) & SVM for European or for certain people, then we can increase the accuracy a bit because these algorithms divide and object into 64 points, we can easily get a good output by training the dataset we get by scanning the left right top to bottom.

* 1. **Rational of the study**

Most of the researchers have been work for Machine Learning (ML) using face detection. But they didn’t find out the proper nationality accuracy of that people, whereas some researcher done this type of work but their outcome didn’t give the whole world 800 millions people outcome, rather than their outcome change one continental to other continental. Moreover, there are several dialects that vary to the areas of the world. People from different regions nationality different face. There is a great lack of nationality detection system in ML.

Image-based applications are playing a significant role in recent period. The applications may be helpful in various ways such as surveys for census, finance, HR and marketing, medical sector, public transportation, solving crimes, corporate office etc. To maximize the uses of image-based applications using nationality dialects for the welfare of the secluded area we get interested to work with this.

* 1. **Research Questions**
* How can we collect the data of different dialects?
* Which methods can be applied for the future extractions?
* Which algorithm should be used for classification to find the best result?
* Can this study be effective for the secluded area inhabitation in the world?
  1. **Expected Outcome**

The prospective result of this research-based project is to develop an application for the general population of the world to know which they belong using the facial length of their image. This type of application may help in many ways. such as:

* This application may help in collecting demographic statistical data such as passport immigration, educational background etc.
* Several kinds of multination company can used this application and also can be categorized by their image and endorsed under trail.
* This research can be beneficial for the foreign security tracking service as they can use it for finding the national of the country by identifying the continental.