Internship Workbook

(Guidelines and Log)

**Course Code:310255**

# (2019 Course)

**Third Year of Computer Engineering**

Year 2023-24

|  |  |  |  |
| --- | --- | --- | --- |
|  | B | 1 | 1 |

**Roll No:** 39053

**Name of the Intern:** Adhish Pawar

**Internship Title**: Deepfake Detection using Hadoop and Pyspark

**Internship Company Details**:

**Name of Internship Supervisor (External)**:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Name of Internship Supervisor (Internal):**

Mrs. Jayaprabha Kanase

A picture containing text, sign

Description automatically generated

# Department of Computer Engineering

# PES’s MODERN COLLEGE OF ENGINEERING, PUNE

**Savitribai Phule Pune University**

**Prologue**

Internships are educational and career development opportunities, providing practical experience in a field or discipline. Internships are far more important as the employers are looking for employees who are properly skilled and having awareness about industry environment, practices, and culture. Internship is structured, short-term, supervised training often focused on task or projects with defined time scales.

Core objective is to expose technical students to the industrial environment, which cannot be simulated/experienced in the classroom and hence creating competent professionals in the industry and to understand the social, economic, and administrative considerations that influence the working environment of industrial organizations.

Engineering internships are intended to provide students with an opportunity to apply conceptual knowledge from academics to the realities of the field work/training. The following guidelines are proposed to give academic credit for the internship undergone as a part of the Third Year Engineering curriculum.

The main purpose of maintaining diary/workbook is to cultivate the habit of documenting. The students should record in the daily training diary the day-to-day account of the observations, impressions, information gathered, and suggestions given, if any. The training diary/workbook should be signed every day by the supervisor.

**General Instructions**

1. Students should enter correct information in the workbook.
2. Get all entries verified by respective guide. No changes are to be made without guide’s permission.
3. Students should report to their respective guides as per the schedule and its log is to be maintained in the workbook.
4. Follow all deadlines and submit all documents strictly as per prescribed formats.
5. The workbook should be produced at the time of all discussions, presentations and examinations.
6. The workbook must be submitted to Internship faculty coordinator/ guide/ department / College after successful examination at the end of year.
7. All documents and reports are to be prepared in given format only (All the formats’ specifications provided adheres to MS Word)
8. Submit hard as well as soft copy and maintain copy with each member.

**PROGRAM EDUCATIONAL OBJECTIVES**

The graduates of Computer Engineering Department will be,

PEO1: Capable of solving real world problems.

PEO2: Capable of working with multidisciplinary projects.

PEO3: Capable to adapt to changing technologies and life management skills.

PEO4: Able to exhibit professional and ethical responsibilities.

**PROGRAM OUTCOMES**

Engineering Graduates will be able to:

1. **Engineering knowledge**: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

2. **Problem analysis**: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

3. **Design/development of solutions**: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

4. **Conduct investigations of complex problems**: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

5. **Modern tool usage**: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

6. **The engineer and society**: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues, and the consequent responsibilities relevant to the professional engineering practice.

7**. Environment and sustainability**: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

8**. Ethics**: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

9. **Individual and teamwork**: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

10. **Communication**: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

11. **Project management and finance**: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one’s own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

12. **Life-long learning**: Recognize the need for and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

### Course Objectives:

Internship provides an excellent opportunity to learner to see how the conceptual aspects learned in classes are integrated into the practical world. Industry/on project experience provides much more professional experience as value addition to classroom teaching.

* To encourage and provide opportunities for students to get professional/personal experience through internships.
* To learn and understand real life/industrial situations.
* To get familiar with various tools and technologies used in industries and their applications.
* To nurture professional and societal ethics.
* To create awareness of social, economic, and administrative considerations in the working environment of industry organizations.

### Course Outcomes:

On completion of the course, learners should be able to

**CO1:** To demonstrate professional competence through industry internship.

**CO2:** To apply knowledge gained through internships to complete academic activities in a professional manner.

**CO3:** To choose appropriate technology and tools to solve given problem.

**CO4:** To demonstrate abilities of a responsible professional and use ethical practices in day-to-day life.

**CO5:** Creating network and social circle and developing relationships with industry people.

**CO6:** To analyze various career opportunities and decide carrier goals.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **The CO-PO Mapping Matrix** | | | | | | | | | | | | |
| **CO/**  **PO** | **PO1** | **PO2** | **PO3** | **PO4** | **PO5** | **PO6** | **PO7** | **PO8** | **PO9** | **PO10** | **PO11** | **PO12** |
| **CO1** | 2 | 2 | 2 | 2 | 3 | 1 | 1 | 1 | 1 | 2 | 1 | 1 |
| **CO2** | 1 | 2 | 2 | 2 | 3 | 2 | 1 | 1 | 1 | 2 | 2 | 1 |
| **CO3** | - | - | - | - | - | 1 | - | - | 2 | 2 | 1 | 1 |
| **CO4** | 2 | - | - | - | - | 2 | 2 | 3 | - | 1 | - | 2 |
| **CO5** | - | - | - | - | - | 1 | 2 | 1 | 1 | 1 | 2 | 1 |
| **CO6** | - | - | - | - | - | 1 | - | - | 2 | 1 | - | 1 |

**Internship Diary/workbook may be evaluated based on the following criteria:**

* Proper and timely documented entries
* Adequacy & quality of information recorded
* Data recorded
* Thought process and recording techniques used
* Organization of the information

**Evaluation through Seminar Presentation/Viva-Voce at the Institute-**

The student will give a seminar based on his training report, before an expert committee constituted by the concerned department as per norms of the institute.

The evaluation will be based on the following criteria:

* Depth of knowledge and skills
* Communication & Presentation Skills
* Teamwork
* Creativity
* Planning & Organizational skills
* Adaptability
* Analytical Skills
* Attitude & Behavior at work
* Societal Understanding
* Ethics
* Regularity and punctuality
* Attendance record
* Diary/Workbook
* Student’s Feedback from External Internship Supervisor

After completion of Internship, the student should prepare a comprehensive report to indicate what he has observed and learnt in the training period.

**STUDENT’S DAILY DIARY/DAILY LOG**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Date** | **18/12/2023** | **Time of Arrival** | **9.30 AM** | **Time of Departure** | **04.30 PM** |
| **Dept./Division** | **TE-B** | **Project Title** | **Deepfake image detection using Apache HADOOP & Apache SPARK.** | | |
| **Main points of the week (Include Figures, if any)** | | | | | |
| Basic research on the topic and limiting the scope of the project. Reviewing various papers for the topics DEEPFAKES, Neural Networks, detection of deepfakes. Getting accustomed to various new terminologies related to the above-mentioned topics.     |  |  |  |  | | --- | --- | --- | --- | | Sr no. | Deepfake Detection technique name/  Neural Networks | Suspicious Text message detection technique name | Datasets used | | 1 | MesoNet | LSTM | faceForensics++ | | 2 | ResNet | BERT | Celeb-DF | | 3 | DenseNet | HMM | DFDC | | 4 | Inception | XGBoost | thispersondoesnotexist.com | | 5 | Xception |  |  | | 6 | MobileNet |  |  | | 7 | VGG |  |  | | 8 | EfficientNet |  |  | | 9 | SqueezeNet |  |  | | 10 | NASNet |  |  | | 11 | PNASNet |  |  |   Reference papers:   * <https://www.journal-dogorangsang.in/no_1_Online_23/14.4_june.pdf> * <https://www.mdpi.com/2624-800X/2/1/7> * <https://www.mdpi.com/2073-8994/14/5/939>   **Comparison Between RestNet and EfficientNet**   |  |  |  | | --- | --- | --- | | **Category** | **ResNet** | **EfficientNet** | | **Efficiency** | While effective, ResNets may become computationally expensive as they get deeper due to the **Linear stacking of layers.** | Focuses on achieving high accuracy with fewer parameters, making it computationally efficient. | | **Depth** | Known for its ability to train very deep networks, reaching hundreds of layers. | Can also be deep, but the emphasis is on balancing depth with width and resolution for efficiency. | | **Scaling** | Typically scales depth **without considering width** and resolution in a unified manner | Utilises a compound scaling method, simultaneously **scaling depth, width, and resolution.** | | **Model Size** | Can be computationally expensive, especially in deeper variants | Designed to achieve a balance between model size and accuracy through efficient scaling |   Split group into 2 for detection of deepfake social media posts and deepfake digital news.  **Number of Parameters**=Filters×Kernel Size×Input Channels×Output Channels   Let's break down how these scaling coefficients affect each dimension:  **Depth (α): The depth of the network is the number of layers**. It is scaled by the α coefficient. Larger α values result in deeper networks, capturing more complex features.  **Width (β): The width of the network is the number of channels (filters) in each laye**r. It is scaled by the β coefficient. Larger β values result in wider layers, allowing the model to capture more features in parallel.  **Resolution (γ): The resolution is the input image size**. It is scaled by the γ coefficient. Larger γ values result in higher-resolution input images, allowing the model to capture more detailed information.     * One of you can study for dataset used/available so far for deepfake, social post analysis   Dataset for implementation:    https://www.kaggle.com/datasets/ciplab/real-and-fake-face-detection | | | | | |

**STUDENT’S DAILY DIARY/DAILY LOG**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Date** | **25/12/2023** | **Time of Arrival** | **9.30 AM** | **Time of Departure** | **04.00 PM** |
| **Dept./Division** | **TE-B** | **Project Title** | **Deepfake image detection using Apache HADOOP & Apache SPARK.** | | |
| **Main points of the week (Include Figures, if any)** | | | | | |
| EfficientNet Accuracy Graph     * Reviewed Efficientnet Architecture and all models of Efficientnet (B0 – B7), Efficientnet B5 was choosen Based on Number of parameters and Accuracy * Studied Compound Scaling Discussed in reference paper given below, Compound Scaling was main idea proposed in Efficientnet Architecture. * Coded Efficientnet (B0-B5)   EfficientNets for DeepFake Detection: Comparison of Pretrained Models    Working of Model with Various Models | | | | | |

**STUDENT’S DAILY DIARY/DAILY LOG**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Date** | **02/01/2024** | **Time of Arrival** | **9.30 AM** | **Time of Departure** | **04.30 PM** |
| **Dept./Division** | **TE-B** | **Project Title** | **Deepfake image detection using Apache HADOOP & Apache SPARK.** | | |
| **Main points of the week (Include Figures, if any)** | | | | | |
| * Studied on Following Concepts in Detail:  1. Inception-ResNet V2 2. LSTM 3. Feature Extraction 4. Temporal Modeling 5. Data Augmentation 6. Adversarial Training 7. Evaluation Metrics: 8. Real-Time Detection   All the concepts you mentioned are related to deep learning approach using Inception-ResNet V2 and LSTM models for effective deepfake detection, showcasing high accuracy and robustness in identifying manipulated digital media. | | | | | |

**STUDENT’S DAILY DIARY/DAILY LOG**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Date** | **08/01/2024** | **Time of Arrival** | **9.30 AM** | **Time of Departure** | **04.00 PM** |
| **Dept./Division** | **TE-B** | **Project Title** | **Deepfake image detection using Apache HADOOP & Apache SPARK.** | | |
| **Main points of the week (Include Figures, if any)** | | | | | |
| Ref: DETECTING DEEPFAKES IN DEEP LEARNING: LEVERAGING INCEPTION-RESNET  V2 AND LSTM MODELS    Key points from the research paper that can be useful for implementing deepfake detection algorithms:  1. Utilization of Deep Learning and CNNs  2. Decision-Level Fusion Techniques  3. Data Augmentation and Mini-Batch Analysis  4. Creation of Multiple CNN Models  5. Cross-Validation for Model Evaluation  6. Use of Publicly Available Datasets  7. Achieving High Classification Accuracy | | | | | |

**STUDENT’S DAILY DIARY/DAILY LOG**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Date** | **15/01/2024** | **Time of Arrival** | **9.30 AM** | **Time of Departure** | **04.00 PM** |
| **Dept./Division** | **TE-B** | **Project Title** | **Deepfake image detection using Apache HADOOP & Apache SPARK.** | | |
| **Main points of the week (Include Figures, if any)** | | | | | |
| * Studied and coded ensemble learning, a concept in deep learning where you can stack multiple small machine learning models to increase the output.     Fig: represents basic idea of ensembling models   * I coded all 5 ensembling techniques using 3 ml models namely, Logistic Regression, Decision Tree Classifier, K Neighbors Classifier * Observation Table  |  |  | | --- | --- | | Type of ensembling | Parameters | | Majority | Accuracy: 85.5% | | Stacking | Accuracy: 98.67% | | Average | MSE: 22% | | Bagging | Accuracy: 99.33% | | Boosting | MSE: 0% | | | | | | |

**STUDENT’S DAILY DIARY/DAILY LOG**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Date** | **22/01/2024** | **Time of Arrival** | **9.30 AM** | **Time of Departure** | **04.00 PM** |
| **Dept./Division** | **TE-B** | **Project Title** | **Deepfake image detection using Apache HADOOP & Apache SPARK.** | | |
| **Main points of the week (Include Figures, if any)** | | | | | |
| * Studied and Coded MBconv2D block which is building block for all efficientnet models * Coded one MBconv layers and got the following outputs * Accuracy: 53.57% * Validation Accuracy: 57.81% * Loss: 63.68% * Validation Loss: 68.84% * Studied the concept of depth wise and pointwise convolution, how both are used together to improve performance.   Fig: MBConv Block    Fig: Efficientnet B0 archi | | | | | |

**STUDENT’S DAILY DIARY/DAILY LOG**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Date** | **29/01/2024** | **Time of Arrival** | **9.30 AM** | **Time of Departure** | **04.00 PM** |
| **Dept./Division** | **TE-B** | **Project Title** | **Deepfake image detection using Apache HADOOP & Apache SPARK.** | | |
| **Main points of the week (Include Figures, if any)** | | | | | |
| A Lightweight ReLU-Based Feature Fusion for Aerial Scene Classification." This study proposes a transfer-learning-based model construction technique called ReLU-Based Feature Fusion (RBFF) for aerial scene classification. The approach extracts feature maps from a pretrained CNN-based model, MobileNetV2, and uses a layer selection strategy based on the characteristics of ReLU activation layers in those blocks. The method utilizes features from the batch normalization layers of selected blocks based on their significance (a measure called β). The extracted features are compressed into a low-dimensional feature space using dimension reduction algorithms like PCA and LDA, then a low-cost SVM classifier is trained for classification | | | | | |

**STUDENT’S DAILY DIARY/DAILY LOG**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Date** | **05/02/2024** | **Time of Arrival** | **9.30 AM** | **Time of Departure** | **04.00 PM** |
| **Dept./Division** | **TE-B** | **Project Title** | **Deepfake image detection using Apache HADOOP & Apache SPARK.** | | |
| **Main points of the week (Include Figures, if any)** | | | | | |
| We have explored several fusion methods for handling cases where the signal length exceeds the designed CNN input length. These methods include voting-based fusion, confidence-based fusion, and feature-based fusion. Simulation results indicate that all three fusion methods surpass the accuracy of the non-fusion method, with confidence-based and feature-based fusion methods performing comparably and outperforming the voting-based approach | | | | | |

**STUDENT’S DAILY DIARY/DAILY LOG**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Date** | **12/02/2024** | **Time of Arrival** | **9.30 AM** | **Time of Departure** | **04.00 PM** |
| **Dept./Division** | **TE-B** | **Project Title** | **Deepfake image detection using Apache HADOOP & Apache SPARK.** | | |
| **Main points of the week (Include Figures, if any)** | | | | | |
| The fusion methods detailed in this paper center on integrating outputs from different layers of the convolutional neural network (CNN) to enhance the accuracy of classification results. In the schematic diagram illustrated in Figure, various segments of the input signal are processed by the CNN, yielding distinct outputs at different layers. By fusing classification results, confidence vectors, or features from these different layers, we implement three fusion methods: voting-based fusion, confidence-based fusion, and feature-based fusion.  Voting-Based Fusion: This method employs the principle of "majority wins" to determine the final classification. Each segment of the input signal is independently processed by the CNN, resulting in classification predictions. These predictions are then combined through a voting mechanism, where the class receiving the most votes is considered the final classification.  Confidence-Based Fusion: In confidence-based fusion, the output of the CNN's softmax layer (a confidence vector) is used for linear fusion. Each confidence vector contains the probabilities that the input signal belongs to each modulation type. The average confidence for each modulation type is calculated across all segments, and the final classification is based on the modulation type with the highest average confidence.  Feature-Based Fusion: This approach utilizes the feature vectors derived from the layer just before the softmax layer of the CNN. By combining the feature vectors from different segments of the input signal, the method applies linear fusion to produce an overall representation of the input. This comprehensive feature set is then used for final classification. | | | | | |

**ATTENDANCE SHEET**

|  |  |
| --- | --- |
| Name of Student | **Adhish Pawar** |
| Roll. No | **39053** |
| Name of Course | **Computer Engineering** |
| Date of Commencement of Internship | **18-12-2023** |
| Date of Completion | **31-1-2024** |
| Organization Name | **PES’s Modern College of Engineering** |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Month & Year** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 |
| **Month & Year** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 1. 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 |

Note:

1. Attendance Sheet should remain affixed in Daily Training Diary. Do not remove or tear it off.

2. Student should sign/initial in the attendance column. Do not mark ‘P’.

3. Holidays should be marked in Red Ink in attendance column. Absent should be marked as ‘A’ in Red Ink.

4. Signature of Company Internship supervisor with company stamp/ seal.

Industry/Internal Supervisor Signature:

Industry/Internal Supervisor Name: Mrs. Jayaprabha Kanase

Email ID: jayaprabha.kanase@moderncoe.edu.in

**Term Work Evaluation Rubrics**

|  |  |  |  |
| --- | --- | --- | --- |
| Internship Workbook(A) | Internship Report(B) | Internship Seminar-Presentation(C) | Total |
| 20 Marks | 30 Marks | 50 Marks | 100 Marks |
|  |  |  |  |

**A) Internship Workbook Evaluation (20 Marks)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Proper and timely documented entries (4M) | Adequacy & quality of information recorded (4M) | Data recorded (4M) | Thought process and recording techniques used(4M) | Organization of the information (4M) | Total  (20 Marks) |
|  |  |  |  |  |  |

**B) Internship Report (30 Marks)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Internship completion certificate (10M) | Scope and Objectives of Internship (5M) | Technological/ Methodological Details(5M) | Results/  Conclusions (5M) | Attendance Record (5M) | Total (30M) |
|  |  |  |  |  |  |

**C) Internship Seminar-Presentation (50 Marks)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Grade (Grade Point)** | **Excellent**  **(10-9)** | **Very Good (6-8)** | **Fair (3-5)** | **Poor (1-2)** |
| **Parameter** |  |  |  |  |
| Depth of knowledge and skills |  |  |  |  |
| Communication & Presentation Skills |  |  |  |  |
| Teamwork |  |  |  |  |
| Creativity |  |  |  |  |
| Planning & Organizational skills |  |  |  |  |
| Adaptability |  |  |  |  |
| Analytical Skills |  |  |  |  |
| Attitude & Behavior at work |  |  |  |  |
| Societal Understanding |  |  |  |  |
| Ethics |  |  |  |  |
| Regularity and punctuality |  |  |  |  |
| Attendance record |  |  |  |  |
| Diary/Workbook |  |  |  |  |
| Student’s Feedback from External Internship Supervisor |  |  |  |  |
| Total |  |  |  |  |
| Out of 50 Marks |  |  |  |  |