National College of Ireland

Improving Public Safety: Advanced Machine Learning for Early Detection of Aggressive Street Dog Behaviours in Asian Urban Environments

MSc Research Project
MSc Data Analytics

Sayan Kumar Ghosh
Student ID: x22211781@student.ncirl.ie

School of Computing
National College of Ireland

Supervisor: Teerath Kumar Menghwar

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: ……. ……Sayan Kumar Ghosh……

Student ID: ……x22211781@student.ncirl.ie……..……

Programme: ……MSc Data Analytics…… **Year:** …2023-2024..

Module: ……MSc Research Project……..……

Supervisor: ……Teerath Kumar Menghwar……..……

Submission Due Date: ……12/08/2024……..……

Project Title: … Improving Public Safety: Advanced Machine Learning for Early Detection of Aggressive Street Dog Behaviours in Asian Urban Environments

Word Count: ……11554…… **Page Count** ……28……..

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ……Sayan Kumar Ghosh……

Date: ……11/08/2024……

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	✓
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	✓
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	✓

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Improving Public Safety: Advanced Machine Learning for Early Detection of Aggressive Street Dog Behaviours in Asian Urban Environments

Sayan Kumar Ghosh
X22211781@student.ncirl.ie

Abstract

This paper introduces a novel multi-model computer vision system for identifying and monitoring the aggression and following behaviours in street dogs of Asia's urban areas. Employing different types of models 2D Convolutional Neural Network, Vision transformer-b16, EfficientNet B3, ResNext -50 and amongst all and sundry, the 2D CNN came out to be very effective, with validation accuracy rate of 98. 12%. This system uses YOLO V5(You Only Look once) for real-time and fast dog detection, and DeepSORT for real-time tracking. It is also a strong factor since the system is compatible with city cameras, making it possible to greatly improve protection in urban environments and perform constant observation accompanied by instant responses, which are essential in complicated urban settings. However, it is confronted with difficulties such as obstacles which it may encounter in passing through, fast motion which a particular object might be displaying, and fluctuations in light situations that sometimes affects its outcomes in terms of brightness and sharpness. Such challenges call for the future improvement, specifically how best to enhance its performance especially in low resolution and dynamic environment scenarios. Through the application of deep learning algorithms, this system quickly scans for attributions of aggression in a dog that could potentially lead to conflicts among people, thus fostering the maintenance of public order. Possible ethical concerns related to the use of this technology to address the problem of stray dogs as well as the impacts of this technology on the urban populations are also discussed. The conclusion of this research work is essential for enhancing interphase relations between humans and dogs in Asian cities. The author also presents the measures applicable for preventing the aggressive behaviour and promoting the cooperation among people and animals.

1 Introduction

With the population of densely populated cities growing, street dogs are becoming a significant threat to the safety of people in Asia, manifested in their aggressive behavior. This issue has now reached danger levels as far as the health of the people in the affected areas is concerned and some intervention measures have to be urgently adopted and implemented. Street dogs roaming act as a major threat to the community by increasing cases of people being attacked by dogs or being infected with diseases like rabies that is a common disease the world over.

1.1 Research Problem

With the observed increase in aggressive behaviours among street dogs the danger of conflicts in urban communities creates a high probability of poses threats to public security; hence the

need to improve the monitoring and management strategies aimed at controlling and reducing conflicts with street dogs. In India there is a problem with the prevalence of dog bite annually at a conservative estimate of 17. 4 million cases annually stated by the Association for the Prevention and control of Rabies in India an NGO based at Bangalore, India. This statistic suggest that there are about one percent of the population of the India of one billion two hundred million. It is now a common fact that at least 4 billion people are bitten by dogs annually. The risk incurred from these bites is the likelihood of contracting rabies a viral disease that impacts the central nervous system this illness is transmitted through the saliva of infected animals such as dogs. The WHO has reported that out of human rabies cases, 99% of transmission occurs from dogs. Shockingly, India has a higher number of rabies victims among all nations; a third of deaths from rabies happen in this country and a large number of the victims are children below fifteen years of age. This highlights the imperative for efficient intervention and prevention mechanisms for handling this public health issue[1]. India has a major problem with regards to the dog population – current pet dogs in India are more than ten million and this figure is set to increase to thirty million in year 2021 while there are more than thirty-five million stray dogs in the country. The animal welfare and control are not well developed in the country for instance breeding of stray dogs and their proliferation. People also fail to care for their pets and abandon them or allow them to go outside on their own, thus getting lost and becoming ill-mannered dogs and joining stray packs. Their aggression presents a very big danger; thus within a year, 2019 to be precise, around 4,146 people were killed and bitten by dogs. Kerala is not one of the top ten states with stray dogs; however, 100 thousand reported dog bite cases occurred in the state in the first seven months of 2022 alone. Evidently, specialists have identified poor garbage disposal, dogs being put out on the streets, and inadequate spaying and neutering or lack of vaccinations for dogs as the principal contributors to this persistent problem. This implies that there is a need to provide a prompt intervention by deploying strategies that will ensure the well being of these dogs and at the same time protect people from being bitten[2].

1.2 Research Question and Objectives

Based on the analysis provided above, a research question for this study has been formulated.

Research Questions

1. **How can the deployment of advanced machine learning models improve the detection and prediction of aggressive behaviors in street dogs compared to traditional monitoring methods?**
2. **What are the technological and societal implications of integrating these systems into urban surveillance networks?**

Hostility in dog especially the street dogs is one of the reasons most frequently mentioned by canine specialists for biting. It is also important to understand that most professionals now

[1] https://scroll.in/article/1056464/why-is-india-seeing-so-much-rage-against-street-dogs
[2] https://timesofindia.indiatimes.com/india/indias-stray-dog-menace-a-crisis-of-proportions/articleshow/99137639.cms

agree that such bites are avoidable. Stray dogs are known to inhabit neighbourhoods alongside human beings and this is not only limited to India but can be observed in most of the developing nations. Usually, they coexist with communities, whether in the countryside or city. Nonetheless, the number of stray dogs roaming on the streets have increased in the recent past. It is associated with changes in the level of food waste caused by the nature of people's incomes, increased populations, rates of urbanization, and the lack of appropriate means of regulating dog overpopulation(Saleem, Jan and Khan, 2020). The advantage of using 2D CNN, Vision Transformer b16, EfficientNet B3, and ResNeXt-50 for early identification of street dogs' aggressive behaviours is their capability of providing accurate and credible insights based on the image analysis. Unlike the conventional approaches that depend on the human eye view or simple technology, each of these models employs deep learning methods in the interpretation of the visuals optimally. For image classification, 2D CNN shows high classification accuracy, while Vision Transformer b16 uses attention mechanisms to capture the spatial relations within the images, Efficient Net B3 takes a middle approach of a good computational cost and high accuracy, and ResNeXt-50 with a very different model architecture provides a strong feature representation. While applying each of the models to the same data set, it is possible to compare their ability to identify various forms of aggressive behaviours, which in turns helps to determine which of the models offers the highest level of accuracy in real time surveillance in urban settings. Difficulties in detecting and recognizing some standardized facial expressions in dogs are caused by a huge variance of their head forms and skin patterns over fur, especially by hypertrophic wrinkles which are genetically inherited and are present on the skin of some dog's breeds permanently. These variations are also making it very hard to define expressions in a uniform way across all types of dogs(Burrows *et al.*, 2021). The compatibility with currently implemented urban surveillance system is one major strength that ensures constant surveillance and immediate intervention in case of further acts of aggression in this system. Information about conversational rates offers early alerts improving security of the public and for that matter the welfare of animals since issues are prevented from escalating. To investigate the behaviour of stray dogs living inside Thai temples towards strangers with the help of a stranger-approach test done between September and November 2018 in five temples. Open-ended field notes were kept on the dogs' behaviours, social status, reaction distances, and their reaction categories like neutrality, withdrawal, threat, vocalization, approach, and tail. Using the stranger-approach test, the initial reactions were observed and noted in the first week only. For the first test, friendly interactions such as feeding and grooming were performed by the observer for approximately 10 minutes once a week for three consecutive weeks and the final test was done in the fourth week. There were improvements seen as the weeks progressed; in the fourth week, majority of the dogs displayed either neutral or approach behaviours and a fewer number demonstrated avoidance or vocalization. The study indicates that dogs that were classified as stray can have positive behavioural changes toward unfamiliar individuals the more they interact. This study also underscores the significance of more research adopting the stranger-approach test with stray dogs(Huo, Kumroptanasan and Na-Lampang, 2019).

2 Related Work

This literature review will continue to discuss early detection of street dog aggression based on previous scholarly papers. In doing so, using bibliographic analysis, the author seeks to comprehend how different approaches have been employed and might be used in the future. It is essential to consider different reviews to get an understanding of the field and to define the current research limitations. This process will assist the author in developing a research question and make a scholarly research with new ideas. It also means that the review will be presented progressively from a more general perspective and then will progress to more focused methodological and result-based arguments. This approach is structured in order to improve knowledge about aggression in street dogs and provide a framework for further research on the issue. The literature review should end in a paragraph that summarises the findings from the state of the art, why the previous solutions are not adequate and justifies the need for your research question.

2.1 Societal Impacts and Public Health Implications

Societal Consequences and Public Health Outcomes in Indian towns and cities, stray dogs are considered by people as their companions in both rural and urban areas in India and have implications for the society as well as health systems. Therefore, what used to be relatively controlled number of stray dogs roaming the street has transformed to a public health issue. When dogs attack people on the streets, the impacts on the individuals may be severe most of the time. This is because they may be attacked by dogs and, sometimes, can even die from the attack. Likewise seeing stray dogs on the streets might make some people feel uncomfortable or even scared, which decreases the overall quality of life of people. If people want to be safe from these dogs, if people want to build healthy societies in India, then one must understand the dynamics of these issues attributed to the particular dogs on streets . Dog bite preventions are not only political and public health problem but also social and economical concern that affect the interest of both the people and the dogs. In this context, specialists in emergency medicine have a tremendous opportunity to do so. Bitten by a dog may cause certain injuries, develop some possibilities of getting infected, and in most cases, result in long-term psychological disorder. In addition to the direct economic impact of the loss, it may also cause anxiety and stress. The purpose of this research is to establish the risk posed by dogs to humans in terms of potential injuries from dog bites and how physicians can avoid that to their patients(Menaka C, 2024). Rabies is one of the deadliest zoonoses that still affects the global public health, and it primarily affects areas of the world where the disease is present. Dogs are one of the main vectors of rabies in the majority of the affected countries, which makes mass immunization of animals one of the effective ways to combat the disease. However, it does not mean that one immunization dosage within the first twelve months of their life is adequate to achieve and maintain the protection antibodies within the booster shot range The studies have proved that the single dosage immunization may not enough to enhance high levels of protection antibodies in animals throughout the year. But, like everything in this world, there is another side of the coin as well. If the animals are administered with several shots of the vaccine over a long period such as during the normal

immunization sessions, the body produces antibodies that stay in the animal's body for a longer time. Therefore, this outcome emphasizes the call to apply booster doses following the first immunization and to have annual immunization programs. It's essential to follow these steps so that even the negative consequences in terms of animates can still be avoided and particularly rabies(Bashir *et al.*, 2022). In order to reduce the growth in the numbers of dogs and to stop the spread of diseases from one dog to the other or from dogs to human beings or where there is conflict between the two, people should castrate such dogs. Sterilization could therefore alter some of the litters and manage their number in the streets particularly of those dogs found in urban settings. This will, in turn, may allow you to possibly get to the early signs of street dog aggression and population control measures such as neutering reducing the problematic influence of the street dog aggression. It is possible to reduce the number of aggressive episodes, which is significant for the security of the residents in a city with the help of sterilization programs of the stray dog population(Chaudhari *et al.*, 2022).

2.2 Exploring the link Between Dog Emotions and Street Dog Aggression

Most of the street dogs found in urban areas reside in countries like India, not in the developed western world. These dogs which are stray and have no owners are important actors in the Indian cities. Some are residents themselves and help other residents with comfort and even moral support. Moreover, in controlling vermin, street dogs contribute to rat eradication, therefore enhancing sanitation in urban settings. Besides, they can also serve as some sorts of security providing the community with information on possible risks or threats. They are however not recognized very formally, but they still have very important functions in the India urban areas proving that dogs are versatile creatures that can function in any environment. Moreover, like other animals, humans also exhibit emotions on their faces as well. However, little prior work has been done precisely on the process of automatic identification of the animal emotional state based on the faces. This is mostly the case because gathering data and having a clear cut knowledge of the animals' emotions is difficult since the animals are unable to speak(Boneh-Shitrit *et al.*, 2022).

2.2.1 Understanding Aggressive Behaviors in Dogs

Dogs also display other behaviours which are genetic and learned from sociological environments. Barking is stimulant response orientated; tail wagging is emotion-indicating where a fast and strong wag is usually a sign of excitement. A dog's barking is done in order to convey something, either to call attention to or to warn. Ear orientation depicts the mood and thus an attentive one is portrayed by an upright one. Olfaction appears to be a component of investigation/interaction most organisms cannot do without. Both possess actions and while they depict bonding, they may also depict demands of attention or intent on playing. Complaining most of the time is to cry of pain, or shout. Laying on their back can mean that they surrender themselves to whoever is giving them a back rub or that they trust someone; Chewing on toys means that the dogs were teething. That these behaviours contribute towards developing sound human-animal relationships is

something many people are unaware of. This in turn is beneficial to the owners as well as their pets, since improved ecological conditions are conducive to healthy evolution. The dog's aggressive behaviour can be identified by the presence of these signs: The behaviours include growling, snarling, display of the teeth, persistent vocalisation, and etc. The changes in body language, which include things like becoming rigid, or having hackles erect, are the signals which are given. The fear-aggression may seem as some protective activities while the territorial-aggression is connected with elements perceived as defense of territories. These concerns assist people to provide patient care, proactive planning and to seek additional help whenever the need arises. That is the reason why, when an owner first notices the sign, safety is guaranteed, and the risk of biting the dog is reduced(De Winkel *et al.*, 2024). For several reasons, it is crucial to understand how dogs feel in order to achieve certain objectives and solve some problems. First of all, we are able to better understand how one can establish deeper relationships with dogs, as illustrated through the example of pets and stray dogs. It can also help us understand what our feeling is when our feeling is similar to that of the dog then we can be in a position to respond as required by the dog and in effect we can be in a position to do what is in the best interest of the dog by providing the necessary care that the dog deserves. Further on, this proves to be useful when identifying signs of aggression/distrust from the dog early enough, which is

paramount in preventing such incidents and ensuring public safety. Since we are now aware that these signs could appear early, it becomes possible for us to address the situation and stop the fight before tragic occurrences that may affect both humans and dogs happen. In conclusion, the study conducted on the stories of the affective worlds of our dog friend improves our knowledge in animal behaviour and enables us to coexist with animals in our society(Ferres, Schloesser and Gloor, 2022). The table 1 indicates various identified themes related to observed behaviours and emotions in dogs.

Table 1(Identified Themes)

Identified Themes	
Theme	**Observed behaviors and/or emotions**
Body posture	Submission, dominance, muscle tension, tail high/low, freeze, etc.
Vocalization behavior	Audible behaviors, such as growling, barking, howling, whining, etc.
Oral behavior	Lip-licking, grooming, yawning, etc.
Observational physiological response to stress	Salivation, urinating, defecation, dilated pupils, trembling, penis protrusion, etc.
Other stress-related behavior	Repetitive behavior, redirected behavior, body shaking, paw lifting, scratching, panting, etc.
Interactions with nonsocial environment	Exploring, hiding, alertness, attention, restlessness, activity, etc.
Expression of emotions	Signs of basic emotions, facial expressions.
Holistic observation of the dog's state	Mental/emotional state: score calculated based on the general picture, taking the context into account.

2.3 Comparing Traditional Techniques and Modern Computer Vision Advances

Before the time of the high technology, the behaviour of humans, regarding dogs, was documented through researching the expertise skills of people of earlier days. The strategy could highlight the whole process of the first humans and dogs' domestication beginning with which days. Over time, humans as direct observers of dog behaviours, obtained

knowledge of what dogs were conveying to each other, and to people. This observation and the subsequent analysis of the activity of the dog dominant and subordinate individuals constituted the foundation of the knowledge about signals, gesture, and the displayed emotions in the canine. The idea was to begin with recognition of such cues which was at the centre of my comprehension. For instance, these signals included the body posture, tail movement, vocal response, and face expressions; by which one would be in a position to understand the intention, temperance, and feeling of the dog. But here, it is only half the problem because human recognition of dog behaviours is also hampered by subjectiveness of people, use of individual knowledge and the lack of a proper method of recording and analysing the tremendous amount of data. The technology is a fluid concept with time, and this ensures that a better and truer machine learning tools can be developed while at the same time exercising the scalability of the tools to incorporate more of the dogs. Likewise the clarity of pain may vary from one breed of dogs to another which may make it harder to identify pain in certain breeds as compared to others. The author proceed to data collection activity by merging data on posture of the dog and the features on its face to give a pain level reading. More specifically, while a number of reports based on canine subjects can be mentioned, reports and investigations conducted on other species, such as, for instance, sheep, mice, pigs and horses are quite examples as well. Thus, the explicit aim of this study is to provide a comprehensive approach to pain assessment in dogs. To achieve this, the author blends the usual approaches as laid down by numerous scholarly journals and other research, and the new strategies proposed by the author. Besides, there was a team of veterinary specialists dedicated to evaluating such videos as the result of pain in dogs. By implementing it, this initiative helps to proceed with a valuable advance in the cases of dogs with pain and our possibility to assist them(Zhu *et al.*, 2023). Moreover, street dogs as they are in a different environment with different peer pressure, different behaviour than pet dogs can be exhibited. Such things as hunger, interaction with other stray dogs or even the environment they are exposed to in big cities affects their behaviours. Street dogs are being highly vigilant, defending their territory and showing pathologic foraging behaviour as they search for food and shelter. It is helpful to acquire this knowledge in assessing their welfare and the issues that they are likely to come across. An assistance for animal motion tracking and pose recognition has helped to change the study of animal behaviour in a way that researchers can sense animal feelings and pain which can be a form of protection for animals. The survey include recognizing pain and emotional states in animals from the facial and body language analysis on the following list; It provides the groundwork by categorizing contemporary initiatives in this space, highlighting challenges, shortcomings in research, and best practice for the movement forward in this regard in this course. Highlighted concepts are affective computing, non-human behaviour analysis, pain estimation, pain recognition, emotion recognition, and computer vision for animals. The questionnaire illustrates the new found realization that animals can not only suffer from pain and depression but they can also experience happiness and thus changed the focus of animal welfare science from the belief that animals can quality of life. The survey is dedicated to the area of computer vision applications for the identification of pain and emotions in animals. Also it acknowledges

the lack of other fields that are related to it such as animal motion tracking, precision livestock farming, methods of landmarks detection and 3D modelling of animal shapes among others. The next section of the document provides the research conducted on the affective states under the mammalian domain and also outlines how computer vision techniques like facial AU can be used to identify the internal affective state of animals(Broomé *et al.*, 2023). A 3D CNN is similar to the neural network for the computer, but is more beneficial as it has a feature of recognizing and analysing those three-dimensional videos. Unlike in one-by-one pics, a 3D CNN can see a full video at once and still catch the motion as well as the transformation occurring. This is why it is not a big challenge for it to follow a pattern and motion in movies. That is part of security cameras to assist in identifying the activities that are not usual or in how doctors use it to take a closer view of the body. In general, it is an intelligent tool that helps computers detect the video scene Computers can make decisions based on deeper features compared to other restrictions imposed on animal behaviour observation, which can be time-consuming to observe them directly, such as sleep, thereby limiting reproduced studies. A 3D CNN is similar to a neural network in computing but is much suitable since it can classify and process videos in three dimensions. This is unlike waiting to make a prediction piecing together one by one pics, a 3D CNN can watch in one sitting of a full video and capture the movement as well as the change. That is why it is relatively easy for it to follow a pattern and motion in movies. It is part of security cameras to assist in noting the activity that is not standard or in an MRI scan to assist doctors in observing the inside of Despite the impressive value of taking sleep as an indication of animal welfare into consideration, adequately, this factor has not been explored scientifically because acquiring data on sleep is very challenging. The overall percent accuracy of the system was calculated to be 89% for onset, duration and fragmentation of sleep in dogs in the real-time observation with the results obtained from implementation of the system. Thus, the automatic system conveyed the impressive view of a useful instrument for conduct and welfare of the animals as the most efficient way of studying the sleep patterns. Evaluation of people based on the video analysis of their behaviours has been enhanced through the incorporation of automated tests and systems that analyse behaviour and which can also enhance the accuracy of the reported behavioural factors. Computerized automated video systems for observation and documentation of animal behaviour are installed within fields of a diverse range of animals: this include both wild animals such as those found in the wild and domestic animals such as those kept for the purpose of farming, rats used in laboratories, bees, fish's among others. However, as has already been noted, automated analysing is yet to drill deep into this area of sleep and, therefore, while the degree to which automated analysing has explored into this area of sleep might still be low, it is beyond any shade of doubt that sleep has been an area of research, which has received a lot of attention in a bid to associate it with healthy and productive lives of humans. In general, this is the wise tool informing computers on scene of videos They can spend much time and make a lot of mistakes to track the animals' behaviours, such as observation of sleep directly, which limits the reproduction of the study. Sleep, the fact that the animal welfare research community is practically always concern with the problem of the measurement, is going to have a great

future potentiality in terms of being a very valuable key parameter in identification of the animal welfare issues. On the other hand, the automated video surveillance which is based on deep learning techniques can be regarded as a possibility to avoid the disadvantages that the measurement of sleep has directly, measuring it through the manual sessions. The analysis of the animal sleep pattern should be automated to achieve the understanding of which statements are true regarding the state and behaviour of animal welfare(Schork *et al.*, 2024). Pet dogs are regarded as beloved friends, and deciphering what dogs' feelings are based on facial cues is beneficial to pet human–dog interactions. This paper presents a work that applies a new CNN model modified by the IWOA to identify dog emotions. Optimizing the choices for this model is know as IWOA-CNN takes into account of those issues attributed to CNN such as slow learning rate, suboptimal solutions and so on. It uses the face detector from the Dlib toolkit to locate the dog's face regions and generates a high-resolution expression set from these images. There is also attached a random dropout layer And L2 regularization can be used to alleviate overfitting and Fine-tune the parameters of the network well. Other classifiers like Support Vector Machine and LeNet-5 also have being compared to IWOA-CNN and have revealed that the latter offer enhanced form of facial expression recognition and therefore, it can be deduced that swarm intelligence algorithms offer a rich means of fine-tuning the parameters of a given model(Mao and Liu, 2023). There are some research papers that use datasets similar to this paper, but they conduct different experiments. Dogs cannot simply respond to the given stimuli and various situations without emotions, as their responses are usually determined by them. The ability to identify dog emotions to a certain extent can be significantly helpful to ethologists and trainers in observance and measurement of dogs with increased accuracy. However, integrating the dog-computer interaction system to be capable of identifying the mood of dogs is advantageous in making these interactions fruitful. The current work presents a novel database including 15,599 images of dogs in different emotional states such as aggressive, anxious, content, or fearful that the authors collected from the Internet. In each picture, a set of taggers with the web-based graphical user interface was used, and all the work was done manually. We have used several avant-garde image classification methods, including an AutoML approach that performed better, resulting in a macro average $f1$ score of 0. The high accuracy of this work is quite remarkable given that these images were obtained directly from the Internet and did not undergo prior processing such as cleaning or segmentation. This methodology provides a novel and efficient way of constructing ISACs in constructivist approaches for building consciousness of dog emotions without the need for aggressive approaches such as physical interventions that could further harm the dogs. Moreover, this proposed approach will generate a vast survey data set to support diligent studies on this significant concern(Hernández-Luquin *et al.*, 2022). According to the proposed surveillance system that has to do with the detection, tracking, and recognition of the emotions of dogs (Canis lupus familiarise), the paper tends to employ sophisticated technologies, significantly the convolutional neural networks (CNNs). The system begins under the YOLOv3 model, which targets on detecting the dogs in every frame of the video. Upon detected, the dogs will be tracked in real time based on the accurately estimator of the real time association metric model accrued as

DeepDogTrack that fused the Kalman filter with CNN. Lastly, people who are informed with specialties in veterinary medicine and dogs weigh on which kinds of dog movement portray anger, happiness, or neutrality. The system after that crops sub-images from the videos to determine if the corresponding images have enough information to identify the dogs' moods. The system will be utilizing LDFDMN which is a LSTM integrated with ResNet for features extracting from the dogs whose emotions will be identified by the LSTM layer. Through trackers and other instruments, the challenges that facial recognition encounters can be solved; however, in the tests practicality the achieved results are rather promising with such a high level of dogs' detection, tracking, and, at the same time, their emotional recognition. By employing the technology, the formation of the system Concerning automatic monitoring of dogs' emotions in operational time is developed to meet the objective of early appearance of a potentially aggressive reaction(Chen *et al.*, 2023).

3 Research Methodology

This section outlines the research methods employed in this study and is done under the Cross-Industry Standard Process for Data Mining (CRISP-DM). As a more rigid framework for solving missions in data mining, CRISP-DM would prove effective in analysing surveillance video data of street dogs in urban areas as per behaviour displaying aggressiveness. This followed a clear and systematic procedure of going through the data to enhance its analysis hence meets the aim of establishing an effective methodology for the analysis of data. All the steps implemented within the CRISP-DM are performed at the high scientific level to ensure the correspondingly high level of data science in order to study the peculiarities of canine conduct in the city context properly. By following CRISP-DM, the author make sure that the project not only would correspond to strict data science procedures but also would contemplate the peculiarities of behaviour analysis in dogs using video surveillance. This structure is beneficial in strengthening and refining the models and techniques wherever necessary in order to attain effective and practical solutions that contribute positively to reducing dangers of cities and uplifting welfare standards for animals.

3.1 Business Understanding

Business Understanding, which forms the first phase of this project, defines the core objectives that have the capacity to act as guidelines for the succeeding phases of a project. The first objective is to design a complex innovative system that is based on machine learning approaches to identify and monitor different aggressive actions of the dogs. It's not just about improving our understanding of canines' behaviours; it is simply about boosting the general safety in cities.

The rationale for this initiative originates from the constant incidences of dog aggression in urban populated areas that present considerable threats to the population's safety and health. With that, the system will be a rich source of information on high-risk patterns of aggression that can prevent the escalation of aggression by implementing the necessary interventions. This precautionary strategy is aimed at helping city planners, animal welfare groups and agencies, and public safety personnel in making sound decision to protect both animal and human lives.

In addition, it is possible to interface this system with other methods of urban management, thus improving the capacity of municipalities to manage street dog populations. It is believed that this integration will increase efficiency and resource allotment and improve the general strategies used to maintain low interaction between humans and dogs(Baslington-Davies *et al.*, 2023). This ideal state is to create the favourable conditions for the coexistence between people and the inhabitants of cities, non threatening environment for both people and animals(Munkeboe *et al.*, 2021). Thus these objectives set right at the start help to specify a direction for technological advancement and to ensure that the technologies that are achieved by society are safe, practical, and applicable to the improvement of urban community living. This synergy is imperative to win the faith of the stakeholders and make a project successful in practical implementation.

3.2 Data Understanding

Data Understanding, the second phase of this project, requires an assessment of the types of data and their source that would be used to build the machine learning models. This phase is very important because most of the subsequent steps of model construction and assessment depend on it.

Static Image Data: The project uses a diverse dataset accessible on Kaggle, which presents numerous images divided by the dogs' emotional states like angry, happy, or sad. These images are very useful when it comes to teaching the models to correctly identify and distinguish the specific emotional indicators. The variation present in this data set enables the models to train from a spectrum that ranges within the extent of canine emotional representations with much emphasis placed on key details concerning different forms of canine expressions.

Video Data: Not only static images to the project consist of video data that portrays dogs in different situations. On one hand, the following video is rather informative in terms of its content and has one primary and two secondary objectives. Firstly, it sets live context that enables us to assess the model in environment that has not been experienced virtually and the environment is as dynamic as in the real world. This aids in determining how accurately behaviour can be tracked and analysed in terms of temporal dynamics which is useful for applications like surveillance in the cities or monitoring animal wellbeing. Secondly, the video data aid in model verification, where they check the ability of the model, trained for static images, to apply its learning on moving images, which is crucial when applying the model in real-life environments.

In analysing the two major types of data sources, we are able to have better understanding of the issues inherent in creating an efficient machine learning system. It is important to ensure that the models that are developed are well grounded technically and also realistic based on the understanding of behaviour recognition in dogs. For one, this phase plays a critical role in laying the groundwork for the subsequent phases of data preparation and model training since the overall strategy is developed and all relevant tasks are defined at this stage.

3.3 Data preparation

The third phase of this project, data preparation is also well-defined in this study where the author systematically transform the acquired image into the best form that would allow correct

analysis by the machine learning models. This encompasses the preparing of the computation environment, the loading and configuration of models, preparation and labelling of data for detection and tracking.

3.3.1 Dog Detection

The dog detection step present in this project is crucial for instance particular targets within image sets and is fundamental for tracking as well as behaviour analysis. Through the YOLOv5 model that is well known for its real-time object detection with high accuracy, this process lets the system detect and localize the dogs in the images gained precision. Contrary to the detection of dogs in the frames this method is designed to give the coordinates of the bounding box that enclose the detected dogs which is critical for tracking and behavioural analysis. Figure 1 represents detection of dog using YOLO v5.

Figure 1(Detection of Dog using YOLOv5)

Detection process involves employing the deep learning model with an objective of seeking through every image to find out whether it has features that defines a dog. Each time a dog is identified, the coordinates, the dimensions of the figure detected, and the degree of certainty are stored. This information is useful as it proves that subsequent steps, for instance, tracking and behaviour analysis are done on accurate data. The reliability of the monitoring system lies with the ability to accurately perceive canines in different settings and under different circumstances including different poses, occlusions, and light conditions which greatly improves the monitoring system's reliability when detecting dogs. Moreover, batch processing helps to handle multiple images at a time rather than encountering them one at a time, and therefore it helps in the speedy detection without compromising the outcome. This step is not just crucial for dealing with a large number of entries but also for making sure that this operation can take place in real time – which is most important in certain high-demand industries.

In general, the detection step paves the way to comprehensive considerations of dog behaviour that is crucial for the achievement of the research objectives of increasing safety in cities, as well as controlling interpersonal relations with dogs in municipalities. Therefore, by guaranteeing high accuracy at this first level, the system creates strong groundwork for intervention and management strategies that may help tackle problems concerning aggressiveness in dogs.

3.3.2 Dog Tracking

The applied tracking step in this project is crucial for investigating the movement and behaviour of dogs within this project, more often once they have been detected. This process tracks the detected dog through the frames of the video footage using the Advanced tracking system, the DeepSort algorithm.

The tracking process in this study starts by initializing the DeepSort tracker with appropriate parameters to ensure the tracking of the dogs' movement. This involves tweaking of certain parameters for a dog for instance, the duration they are captured in the frame of video and ensure that the detections should not be overlapping to avoid a repeated detection of the same dog within video frame. If a dog is located its position and detection quality is noted giving a point to start with on how to track the dog. Upon reception of new images, the tracker adjusts the tracks of each dog and refines tracking to estimate each dog's movement over the frames. For every frame, the coordinates of the bounding box of each dog along with a unique tracking id are provided in order to differentiate between multiple dogs and track their respective movements.

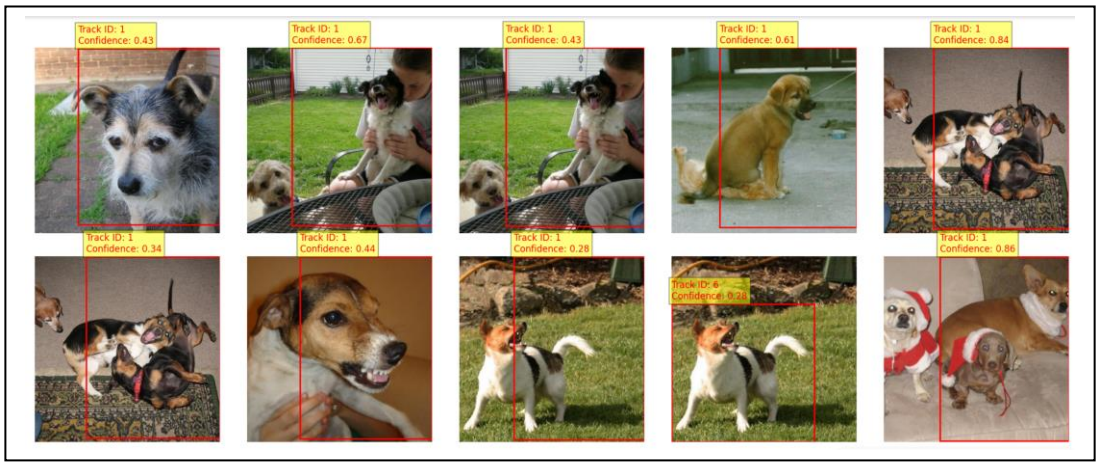


Figure 2(Dog Tracking)

Last but not least, all the above tracking information is stored in a structured JSON file and contains daily data on each dog's location preferences and their corresponding states, for example, 'angry,' 'sad,' or 'happy. 'This comprehensive tracking setup is important for building the picture of the behaviour of dogs in urban contexts and improves the study's capacity to make observations. Figure 2 represents that how the dog is being tracked by assigning track ID and confidence.

3.3.3 Data Preprocessing

The preprocessing step is crucial in this project mainly because it helps in the preparation of the tracked dog images that will be subject to machine learning analysis. This phase entails the process of converting the images to a form that is most appropriate in deep learning models with emphasis on enhancing the images to the input standard of the CNNs.

Image preprocessing is initiated by importing tracked results from a JSON-file containing coordinated description of each dog's behaviour, which is necessary for mapping images to behaviour. Resizing is done to change images to 224 * 224 pixels, change cv2 format to PIL

format, change them to tensors and then to normalize them for the efficient running of the model and also improve its prediction rates. All these transformations are done in parallel with the help of a joblib which helps in faster computation and uniform processing of images.
Further preprocessing of images, involves partitioning images by emotions such as anger, sadness, happiness among others and assigning integer values that would be easier for machine learning algorithm to identify. The dataset itself is subsequently divided into the training set and the validation set depending on the model developed to enhance its accuracy and avoid overfitting. Further, the datasets are stored in the PyTorch format to make it simple to access the data during the training phase and also allow the flexibility of reusing the data at different points in other models or even in different workshops, which increases the possibility of achieving the goal of research accomplishment and reproducibility.

3.4 Modelling

The fourth phase of this project, Modelling, is the interplay of the theory work of machine learning and real-world implementation. This phase covers the actual process of building and training of various deep learning models that can be used to identify the emotional state of dogs. These are trained using image datasets and then benchmarked on videos ensuring their efficiency within dynamical realistic environments.

During this important step of the selection of the model for the project, we analysed several models to define which approach would be the most suitable for the identification of the emotions in dogs based on images. This process includes the usage of various models trained on our dataset in order to determine which would give the highest results in terms of the predetermined indicators. First, the data is thoroughly analysed, and the research question related to the emotions of dogs – whether they are angry, sad, or happy – is discussed in detail; subsequently, it is placed into the multi-class classification context. The applicable models for this classification task are the deep learning models that have been proved to be useful in image processing. Based on the previous studies and given the fact that our data set is massive and complex, the chosen models include 2D CNN, EfficientNet B3, ResNeXt-50, and Vision Transformer (ViT). Every single model is tuned to different hyperparameters with the aim of enhancing the performance when it comes to the detection of emotion in different types of canines within the urban areas. This systematic manner guarantees that the selected model meets our data's requirements not only from the perspective of data but also from the general strive for improving the safety of cities with the help of understanding canine behaviour.

4 Design Specification

The design specification can be described as a set of guidelines and characteristics that define our project: the requirements, limitations, and objectives of our machine learning system which is designed to recognize and classify dog emotions. It is the critical phase in project management that highlights the structures, strategies, and expected results of algorithms to be used.

4.1 Modelling Technique

Vision Transformer (ViT) b16 - In this project, the vision transformer b16 model which is famously known for detail image recognition was used in recognizing the emotions of the dogs using pre processed images. The model was trained on the set of images with specific emotions, for example, anger, sadness, and happiness. Every picture was resized and then enhanced, as well as, normalized for the purpose of maintaining a better training progress. This training consisted of 20 epochs during which the model's task was to differentiate between various emotional conditions of the dogs using the extracted characteristics of the pictures.

The beginning of the process was done using pre-processed image data that were in training and validation data setting. The ViT model was then applied to the images in batches, and the model parameters were tuned by assessing performance on the validation set in the aim of avoiding overfitting and attaining a high level of accuracy. The results of the training session revealed progressive increase in the accuracy of identification of the dog emotions. The validation accuracy, although marginally higher than training accuracy, was significantly lower which indicated that the model has poor ability to extrapolate the results obtained from training to a new unseen images. This indicates that the parameters of the model should be optimized again, or the training data set should be expanded to make the model more predictive.

ResNeXt-50 - In this project ResNeXt-50 model was used which is combined model of ResNet and Inception for the purpose of classifying emotions in dogs from images. This kind of model can effectively manage the computational resources and acquire various features because of its structure. The workflow involved feeding pre-processed images that were prelabelled with emotions such as happy, sad and angry among others. The values of these images were also normalized so that the images were all of the same scale and the model could perform better. To decide the emotions that a dog is showing, the ResNeXt-50 model was trained on 10 training epochs. The training process ensured the flow control by the model reweighting over the performance figures determined against the validation set to avoid the cases of overtraining and the model applicability to the new data.

During each epoch the criteria like loss and accuracy were measured and gradually it was seen that the model was learning to classify emotions correctly. Based on the final validation results, the ResNeXt-50 model proved to be efficient, and the final percentage of correct predictions was more than 93%.

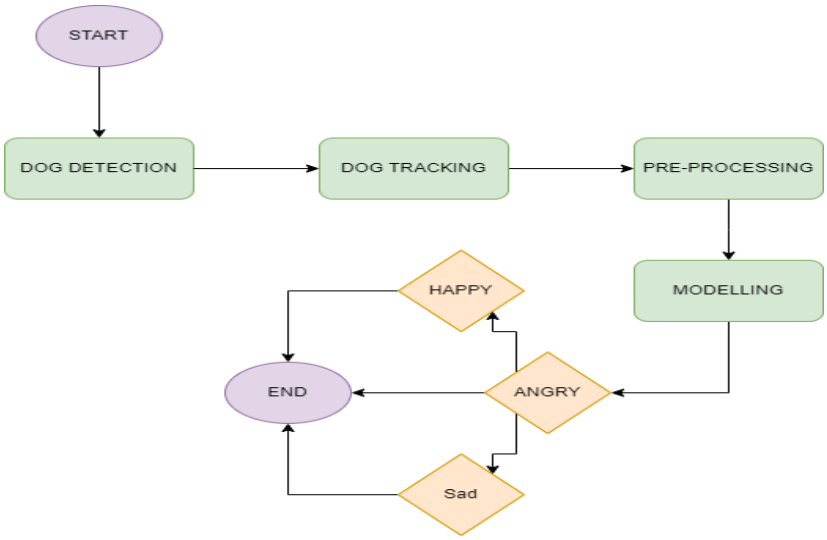
Each epoch, the trained model was stored which can be useful in the future or for further fine-tuning of the model. This training sequence also emphasizes why it is essential to use such a complicated structure of the neural network as ResNeXt-50 for such complicated tasks as the emotion recognition in dogs.

EfficientNet B3 - For this work, the EfficientNet B3 model, which is regarded both lighter and accurate, was employed to classify the emotions in dogs from the images. It is advantageous in computer architecture since it helps in saving computation ingredients while at the same time offering high computation rates, particularly for image data.

This was followed by loading and scaling pre-processed images to be ready for the model to take them in. For training and validation, over 10 epochs, the model was trained to differentiate between a happy, sad or angry dog among other emotions.

When evaluating the model during training, it was observed that the accuracy increased considerably implying the model's ability to learn from the data. After the training session, the model even spiked validation accuracy of 95.23%, which is great for generalization to new, unseen photographs. The training epochs' performances were documented and illustrated to reveal consistent development in the loss and accuracy. At the end of each epoch, the model was then saved, to recall the best performing model later on. Thus, the positive result of the EfficientNet B3 model proves that this model could be used in real-life situations, including the recognition and analysis of dog emotions effectively, which, in turn, will be useful in various scenarios related to the understanding and prediction of animal behaviour.

2D CNN - This study showed that a 2D Convolutional Neural Network (CNN) effectively flagged dog emotions from images and achieved fast training with the results given in five iterations only. Initially the model was getting trained with moderate accuracy but in the 5th epoch the accuracy started boosting with up to 98%. This rapid learning exemplify the flexibility of the model in terms of its capacity to process and learn from raw visual data, which is vital for sensing different feelings such as or anger, depression, or joy. The training and validation phases were used to postpone the test phase and compare the results to prove that the model could generalize rather than memorizing information. Such high performance substantiates the capability of the 2D CNN for application in those scenarios that need a precise identification of different behaviours of the dog; in other words, it extends the capacity to comprehend animals and facilitate the human-animal interface through automated method of analysing visual data. This structure emphasizes the possibility and feasibility of integrating a fast and efficient system of neural networks for emotion recognition in canines.



Flowchart of Research Project

4.2 Evaluation Technique

In this project, assessing the performance of the trained machine learning algorithms used in the identification of emotions experienced by dogs from images entails the consideration of the following evaluation metrics, which is a set of parameters used when measuring the efficiency of the implemented models.

1. Accuracy : This metric depicts the ratio of correctly predicted instances to the total predicted instances. Accuracy based on the number of correct predictions and total predictions provides a snapshot view of model performance. Nevertheless, it does not always give an overall picture of the contingency due to the unequal distribution of data in different categories.

2. Confusion Matrix: It is a table format that assists in portraying the performance of the classification model. It provides the count of false and true predictions along with the respective class. This matrix is especially informative when it comes to assessing the model's performance concerning specific classes and recognizing any potential class imbalance.

3. Precision and Recall: They are useful for performance evaluation of models where false positive and false negatives ratio is of value:

- Precision measures how many of the positively predicted instances are actually positive. Defined as the ratio of accurately detected positive cases to the total of cases that are correctly predicted as positive and those misdiagnosed as such, it indicates the extent of confidence that can be placed in the model's ability to accurately identify positive cases.
- Recall (or sensitivity) shows the model's capacity to pick out all the samples of interest in a given dataset. It is defined as a ratio of true positive genuine impostors in the resultant model to the total of true positives and false negatives; namely, it measures how effectively the model is capable capturing as many positives as possible.

4. F1 Score: This is a better measure since it weighs both precision and recall and attaches a single value to both of them, the harmonic mean. The F1 score is particularly valuable if the quality of the results depends on the combination of precision and recall; if the distribution of data from different classes is uneven. Find it more suitable than accuracy, especially when dealing with class-imbalanced datasets.

5. Train Loss: Measures how well the model is fitting the training data.

6. Test Loss: Used as the last parameter for indicating how effectively the model is going to predict on unseen data which will determine its applicability in real-world scenarios.

7. Validation Loss: Used in tuning the model's hyperparameters while at the same time avoiding the use of the test data. It is helpful in the sense of not allowing the model to over-fit since the data it uses is different from the training data.

5 Implementation

In the last exponent of the process of implementing of the solution of dog emotions recognition in images, some major components was designed and incorporated to constitute the whole

system. In this part of the study, information about the deliverables generated, the instruments applied and a general approach employed in the completion of the project is provided.

5.1 Tools and Languages

Google Colab: The whole project was done in Google Colab to make the best use of its GPUs and the real-time collaboration feature which boosted the model training time.
Python: Python was the primary language of choice because of its availability in the abundance of libraries that are related to data manipulation and machine learning.
PyTorch and torchvision: These libraries were used to train the neural network models as well as to build the same. PyTorch was used to define and manage the models and torchvision for the standard models and image transformations.
Matplotlib: It needs to mention that this library was used for the purpose of visualization of the data and results whereby such as the training loss and validation loss, the training accuracy and the validation accuracy throughout the epochs.
OpenCV: Used in the pre-processing step for image processing such as resizing and formatting of the images to be used for analysis.

5.2 Output

Transformed Data:
To do this the author pre-processed the raw images and converted them to the proper size and format for input into our machine learning algorithms. Other operations included reshaping the images, standardizing the pixel intensities as well as the casting of images to tensors.
Models Developed:
For primary experiments, the models used were 2D CNNs, Efficient Net B3, Vision Transformer B16, and ResNeXt-50. These models were employed for the purpose of detecting as well as analysing the feelings that were embodied by dogs in the respective images.
Code Written:
Great scripts were written to pre-process the data, train, validate and test the models. These scripts also ensured the orderly transfer of data from the pre-processing phase to the production of the model.
The implementation was organized more systematically: first, all the images to be inserted needed to be pre-processed to ensure the corresponded to deep learning requirements. Subsequently, various new models were built and tested with the help of transformed set while constantly checking the performance in terms of each model's effectiveness. The selected model is such that it had the highest validation accuracy and the lowest validation loss. The final stage of the implementation phase involved the last setup of the final built-in model to check on ready to deploy structure to analyse the new images and classify the emotion of the dogs precisely. This configuration offers a solid base for a subsequent work and real-life utilization, for example, tracking the state of pets and other stray animals in cities.

6 Evaluation

This section contains a comprehensive analysis of the experiments that were carried during two phases of the project. This paper aims at reviewing the results of the study and identifying the suitability of different research techniques as well as their implications to real life scenarios and to the academic fields. Due to its specificity and pertinence to the study aims, the analysis concentrates on the findings most relevant to the objectives and employs statistical calculation and diagrams for the thorough examination of the results and averages.

6.1 Final Approach

In this project, the author tried to evaluate different models' performances, namely 2D CNN, ResNeXt-50, EfficientNet B3, and Vision Transformer (ViT). This method was supposed to compare the intrinsic strengths of each model when it comes to the analysis of raw data.

ResNeXt-50: The author pointing that in a beginning of tests faced issue with stability with the lower accuracy and losses figures. But later on it has been enhanced considerably and the model got a training accuracy of almost 97 percent and validation accuracy of nearly 92 percent. One possible cause of this is that there might be some sensitivity of the model to certain data properties that needs further modification or in which the introduction of more augmented data could be beneficial.

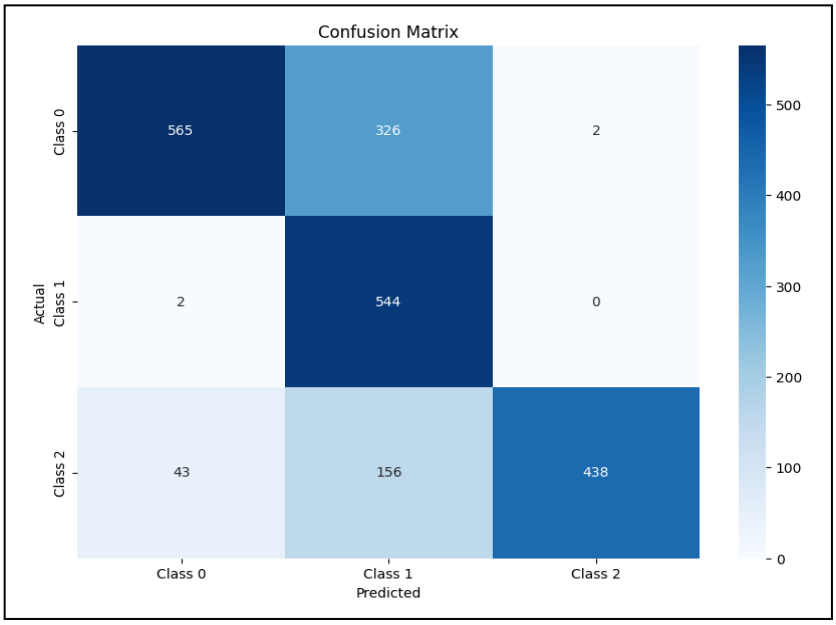


Figure 3(Confusion Matrix of ResNeXt-50)

EfficientNet B3: Started with a great first epoch and quickly and consistently increased performance in terms of all parameters and reached training accuracy of more than 94% and close to 93% of validation accuracy. A high precision and recall therefore clearly shows the efficiency of the algorithm in classifying the different classes generated in the dataset and the reduction in loss over the epochs further proves the efficiency of the model.

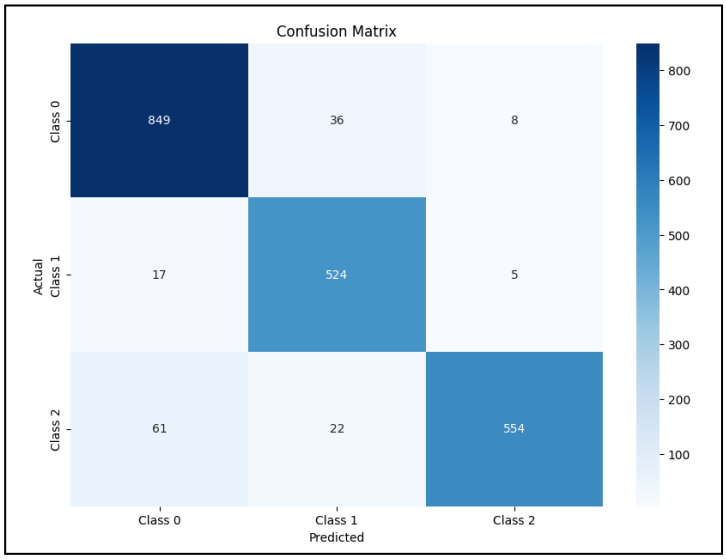


Figure 4(Confusion Matrix of EfficientNet B3)

Vision Transformer (ViT): When using the training dataset, it achieved an accuracy of approximately 37.23% in training and about 40% at the end. But training and validation accuracies were under 44% and that showed that ViT could not cope well with the dataset without augmentation. This implies that ViT might need more training data, more epochs, or even tuning of hyperparameters in order to achieve the best performance in such scenarios.

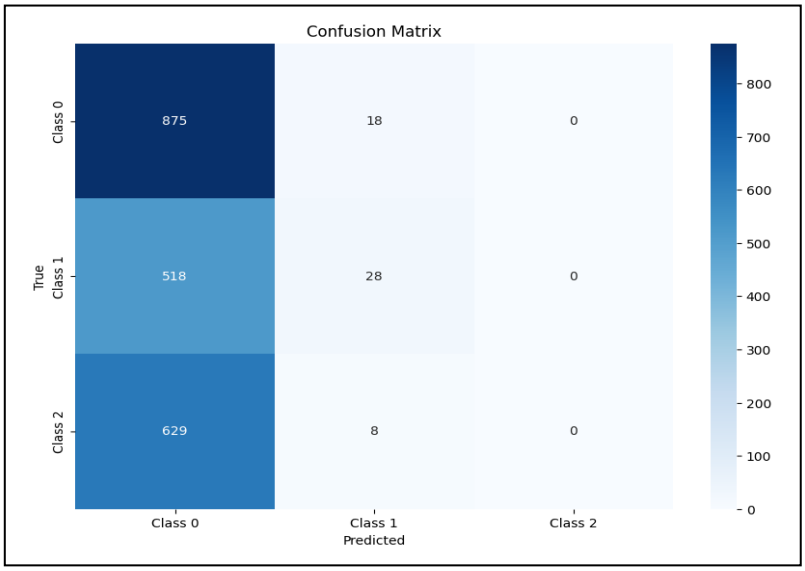


Figure 5(Confusion Matrix of Vision Transformer b16)

2D CNN: Demonstrate outstanding ability to adapt and learn, ending with accuracy of training near 99% and the validation at about 98%. In light of the obtained values for precision, recall, and F1, it can be concluded to have a high generalization potential in the dataset. An average loss of 0.0003 is further proof that the proposed model is well suited for extracting appropriate features from the dataset without the use of data augmentation.

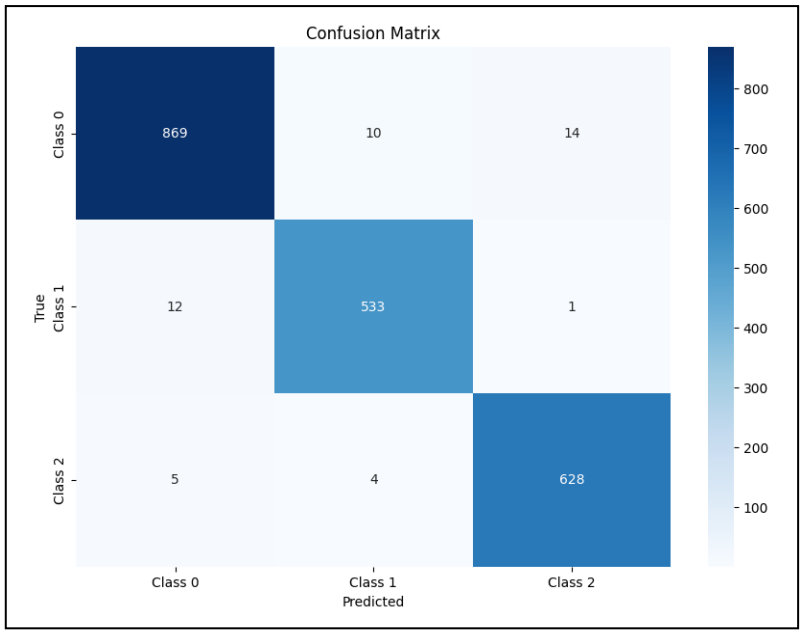


Figure 6(Confusion Matrix of 2D CNN)

Table 2(Results of all Models applied)

Model Name	Epochs	Training Accuracy	Validation Accuracy	Training Loss	Validation Loss	F1 Score
2D CNN	5	98.83%	97.78%	0.0510	0.1527	0.9778
ResNeXt 50	10	97.43%	74.52%	0.0830	0.8406	0.7524
EfficientNet B3	10	94.65%	92.82%	0.1496	0.2041	0.9274
Vision Transformer (ViT)	20	40.39%	43.83%	1.0846	1.0695	0.2312

Overview

The results justified the 2D CNN model training which was seen to have a higher accurate predictability closer to optimal values for all five metrics that were deployed; accuracy, precision, recall, and F1 by the time the training process was complete. It had a very fast learning rate and was also quite capable of predicting, even when it was trained on the raw data, which was a plus for the model since this was something that other models could not accomplish. In particular, the accuracy of ResNeXt-50 either increased or decreased depending on the particularities of the given problem, which indicates that this net can be improved even further through some additional fine-tuning. Although it started at a lesser level, it increased to levels of reasonable accuracy, thus proving how it develops specialized features after the first problem of adaptability.

Therefore, while the performance of EfficientNet B3 regarding the accuracy and the loss rate appeared to be acceptable and even constantly improving within the iteration, it was ranked in the second tier but not very distant from the first one. This layer's scores are consequently high,

and it is only applied to problems where reliable extrapolations of inputs with substantial variation are required.

Vision Transformer (ViT) was not as suitable for the dataset when no further changes were made to the model. This model had a relatively low accuracy, and it could initially mean that it requires more data or computational resources to be trained as effectively as the other models.

Best Model Selection

Based on the evaluation, 2D CNN is identified as the best model for this particular dataset. Among the tested algorithms, the 2D Convolutional Neural Network (CNN) had the highest scores for a number of rather obvious reasons. First, it yielded the highest accuracy in the shortest time and reliably exhibited high accuracy with distinct validation tests that prove its stable learning and good old performance on the new data. Second, the model was equally strong in evaluating raw data hence unveiling its inherent characteristic of feature detection and learning of profound information without pre-processing the input data. In addition, that is why 2D CNN was among the models that took the shortest period of time in training while being computationally-efficient. It is fast, accurate and its optimized implementation makes it suitable to be used in real world applications where performance and efficiency are both key factors. Figure 7 represents accuracy and loss of both training and validation. After testing the 2D CNN model on images and a video the results are shown in figure 8 and figure 9 respectively.

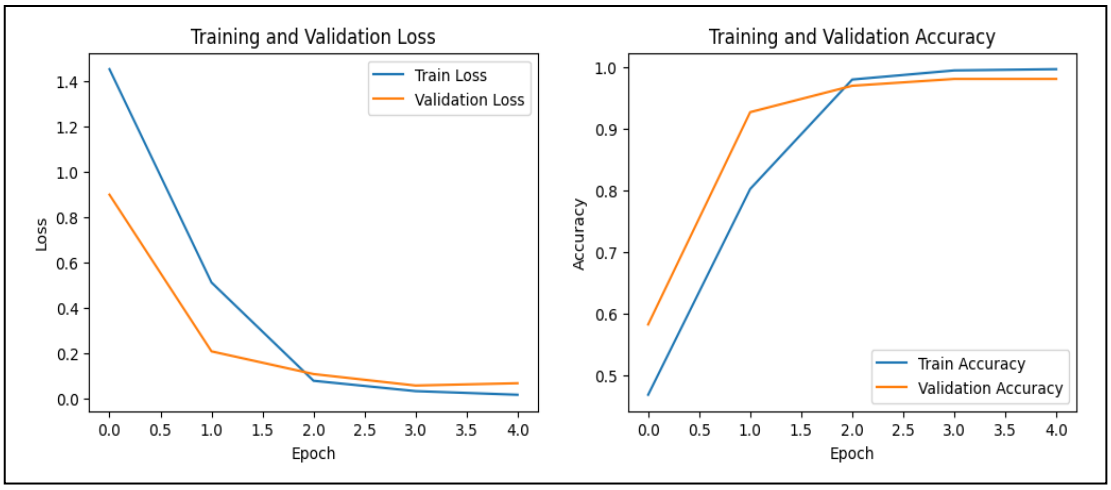


Figure 7(2D CNN Validation Loss and Validation Accuracy)

Figure 8(Dog Emotion Prediction on Images)

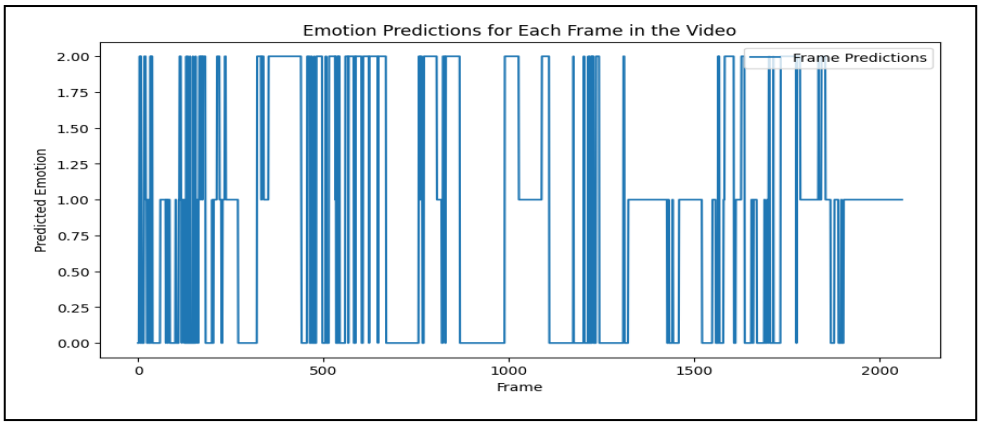


Figure 9(Dog Emotion Detection on a Video)

6.2 Discussion

The research done through various ways using different models has proved to be comprehensive and has given the following findings. As with every model, each model type had its outstanding features and limitations, thus confirming the necessity of the right choice of the model depending on the particular characteristics of the dataset and the goals of the project.

Critique of the Experiments:

The method of analysis undertaken through various ways using different models has provided sufficient analysis results. The selected models highlighted the advantages and the drawbacks accompanying their application, thus, highlighting that model choice depends on dataset properties and the project goals.

Suggestions for Improvement:

Some of the measures that could be employed to increase the efficiency of the models employed in this research include the following. When training these models adding more epochs or greater advances in early stopping may provide the models with more chances to

learn and improve possibly increasing the accuracy and efficiency of the models. Some approaches which may yield improved results include utilizing various kinds of models in ensembles or in constructing a hybrid architecture, since the best characteristic of every specific model could be used effectively for the general improvement of the performance. Expanding on the activities above, more reliable tactics in validation such as using the k-fold cross-validation method would be beneficial as it provides a sound standard of check on the model results in different data segments. The reader may wonder why is not using a more diverse set of images or why are not using transfer learning from a model pre trained with bigger data set, the reply will be simple, the models are able to learn very well from the given data set but they lack the capability to generalize what they learned from the data set to the real world.

7 Conclusion and Future Work

The main purpose of this study was to assess the performance of contemporary deep learning architectures in detecting dogs' feelings based on images without data augmentation. The study tested four models: including 2D Convolutional Neural Networks (CNNs), Vision Transformer (ViT), EfficientNet B3, and ResNeXt-50. The best model was the 2D CNN with the most accurate validation accuracy and F1 score for the selected classes. Vision Transformer and other large models demonstrated good results but needed more data to be trained to achieve their potential; thus, they had a higher requirement for large datasets. Also, non-augmented performance of EfficientNet B3 and ResNeXt-50 were also quite impressive indicating that these can be used in real world scenarios where the systems have to be resource effective and scalable. This study identifies which models would be most effective for low-data image-based emotion recognition and supports the idea that data quality and especially data variety represent an essential prerequisite to the success of the models. Nevertheless, there is a flaw in the use of a limited data set and in not applying data augmentation to the study; they also could not carry out a vast number of experiments due to the issues in computation.

Future Work:
Several directions for the further development of the models of recognition of emotions with regards to pets can be suggested for the future research to increase the effectiveness and usability of such models. One of the studies relates to the integration of the highest performing models with IoT devices to detect emotions in real-time; this can be further sold in the pet care niche. Another avenue is the integration of multimodal data, which comprises of, audio, video, and physiological data for the enhancement of the efficacy and the reliability of the emotion recognition systems. It is also another thought that is possible to develop models that are able to forecast future emotions or probable misbehaviour on the basis of data collected in the past that would give a preventive tool for any owner of a pet and veterinarian. However, further it also be important to also address the ethical perspective and infringement of rights to privacy to contemplate if such systems would always respect the welfare of the pets. These steps are intended to improve present day models and extend scholarship into potentially more organismic processes that may prove beneficial in both educational institutions and pet care industries.

References

Bashir, K. *et al.* (2022) 'The Lived-In Experience and Psychological Recount of Dog Bite Victims Visiting the Anti-Rabies Clinic in Kashmir: A Qualitative Study', *Healthline*, 13(3), pp. 208–213. Available at: https://doi.org/10.51957/Healthline_407_2022.

Baslington-Davies, A. *et al.* (2023) 'An Assessment of Scientific Evidence Relating to the Effect of Early Experience on the Risk of Human-Directed Aggression by Adult Dogs', *Animals*, 13(14), p. 2329. Available at: https://doi.org/10.3390/ani13142329.

Boneh-Shitrit, T. *et al.* (2022) 'Deep Learning Models for Automated Classification of Dog Emotional States from Facial Expressions'. arXiv. Available at: https://doi.org/10.48550/ARXIV.2206.05619.

Broomé, S. *et al.* (2023) 'Going Deeper than Tracking: A Survey of Computer-Vision Based Recognition of Animal Pain and Emotions', *International Journal of Computer Vision*, 131(2), pp. 572–590. Available at: https://doi.org/10.1007/s11263-022-01716-3.

Burrows, A.M. *et al.* (2021) 'Dog faces exhibit anatomical differences in comparison to other domestic animals', *The Anatomical Record*, 304(1), pp. 231–241. Available at: https://doi.org/10.1002/ar.24507.

Chaudhari, A. *et al.* (2022) 'Technology for Improving Street Dog Welfare and Capturing Data in Digital Format during Street Dog Sterilisation Programmes', *Animals*, 12(15), p. 2000. Available at: https://doi.org/10.3390/ani12152000.

Chen, H.-Y. *et al.* (2023) 'Convolutional Neural Network-Based Automated System for Dog Tracking and Emotion Recognition in Video Surveillance', *Applied Sciences*, 13(7), p. 4596. Available at: https://doi.org/10.3390/app13074596.

De Winkel, T. *et al.* (2024) 'Observational behaviors and emotions to assess welfare of dogs: A systematic review', *Journal of Veterinary Behavior*, 72, pp. 1–17. Available at: https://doi.org/10.1016/j.jveb.2023.12.007.

Ferres, K., Schloesser, T. and Gloor, P.A. (2022) 'Predicting Dog Emotions Based on Posture Analysis Using DeepLabCut', *Future Internet*, 14(4), p. 97. Available at: https://doi.org/10.3390/fi14040097.

Hernández-Luquin, F. *et al.* (2022) 'Dog emotion recognition from images in the wild: DEBIw dataset and first results', in *Proceedings of the Ninth International Conference on Animal-Computer Interaction*. *ACI'22: Ninth International Conference on Animal-Computer Interaction*, Newcastle-upon-Tyne United Kingdom: ACM, pp. 1–13. Available at: https://doi.org/10.1145/3565995.3566041.

Huo, X., Kumroptanasan, P. and Na-Lampang, P. (2019) 'A behavioral study of stray dogs in temples in Nakhon Ratchasima municipality, Thailand', *The Thai Journal of Veterinary Medicine*, 49(4), pp. 421–425. Available at: https://doi.org/10.56808/2985-1130.3008.

Mao, Y. and Liu, Y. (2023) 'Pet dog facial expression recognition based on convolutional neural network and improved whale optimization algorithm', *Scientific Reports*, 13(1), p. 3314. Available at: https://doi.org/10.1038/s41598-023-30442-0.

Menaka C, P.G., Anuradha Rohinkar, (2024) ‘Analysis of Dog Bite Injuries to Humans: Strategies for Prevention by Emergency Physicians’, *Revista Electronica de Veterinaria,* 25(1), pp. 1–14.

Munkeboe, N. *et al.* (2021) ‘Comparing Behavioural Problems in Imported Street Dogs and Domestically Reared Danish Dogs—The Views of Dog Owners and Veterinarians’, *Animals*, 11(5), p. 1436. Available at: https://doi.org/10.3390/ani11051436.

Saleem, S., Jan, S. and Khan, Sm.S. (2020) ‘Qualitative analysis of the perception of street dog bite victims and implication for the prevention of dog bites at a teaching hospital anti-rabies Clinic’, *Journal of Family Medicine and Primary Care*, 9(8), p. 4118. Available at: https://doi.org/10.4103/jfmpc.jfmpc_522_20.

Schork, I. *et al.* (2024) ‘Automated Observations of Dogs’ Resting Behaviour Patterns Using Artificial Intelligence and Their Similarity to Behavioural Observations’, *Animals*, 14(7), p. 1109. Available at: https://doi.org/10.3390/ani14071109.

Zhu, H. *et al.* (2023) ‘Video-based estimation of pain indicators in dogs’, in *2023 11th International Conference on Affective Computing and Intelligent Interaction (ACII)*. *2023 11th International Conference on Affective Computing and Intelligent Interaction (ACII)*, Cambridge, MA, USA: IEEE, pp. 1–8. Available at: https://doi.org/10.1109/ACII59096.2023.10388142.