IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

GALACTICAL IMAGE DETECTION AND IDENTIFICATION - A SURVEY REPORT

Dr. Prashantha H S
Professor, Department of Computer Science and Engineering
KS Institute of Technology
Bengaluru, Karnataka

Mr. Najeeb Muthaheed Arin ul Haq
UG Student, Department of Computer Science and Engineering
KS Institute of Technology
Bengaluru, Karnataka

Mr. R Kushal Sai
UG Student, Department of Computer Science and Engineering
KS Institute of Technology
Bengaluru, Karnataka

Mr. Talib Mukhtar Choda
UG Student, Department of Computer Science and Engineering
KS Institute of Technology
Bengaluru, Karnataka

Mr. Rakshith C Mahaladkar
UG Student, Department of Computer Science and Engineering
KS Institute of Technology
Bengaluru, Karnataka

Abstract The human visual system effortlessly recognizes and identifies objects within its field of view, showcasing remarkable speed and accuracy in tasks such as object detection and identification. Valuable clues regarding the origin and evolution of the universe are intricately woven into the shapes and formations of galaxies. However, automating the classification of galaxies from their images poses challenges due to the faintness of the galaxy images, interference from bright background stars, and inherent image noise. To address these complexities, we propose a method for the automatic detection and classification of galaxies. This method incorporates a data augmentation procedure,enhancing the robustness of trained models against data variations arising from different instruments and contrast-stretching functions. This innovative approach is a pivotal component of an expanding open-source computer vision repository dedicated to processing and analyzing extensive galactical datasets. The repository integrates high-performance deep learning algorithms, specifically leveraging the You Only Look Once (YOLO) model and other advanced techniques in image processing and computer vision. The underlying model, trained through deep learning methodologies, exemplifies the intersection of cutting-edge technology and the exploration of the cosmos.

Keywords – Galaxy Identification, You Look Only Once (YOLO), Deep Learning, Image Recognition.

I. INTRODUCTION

In the early 1900s, astronomers found galaxies, marking a big moment in space exploration. Today, astronomy is full of data. By studying different types of galaxies, scientists learn important details about how our universe began and grew. The data we collect about galaxies keeps getting bigger and more complicated. We can even estimate how old galaxies are and what their past was like by lookingat how they move, their shapes, and the chemicals they're made of. The shapes of galaxies are a key to understanding how our universe formed and changed over time. To make this easier, scientists are focusing on identifying galaxies. In recent times, smart computer techniques, like deep learning, have become really useful in areas with lots of data, such as genetics, high-energy physics, and astronomy. Inastronomy, these techniques help find unusual galaxies. The whole process involves four steps: 1. Gather and organize the galaxy data. 2. Use the YOLO system. 3. Train the YOLO system to recognize galaxies. 4. Test how well the trained system works.

Image detection, often referred to as image recognition or computer vision, is a technology that enables machines to interpret and understand the content within images. This technology has various applications, ranging from facial recognition to galactical image detection. Image detection involves the use of algorithms and machine learning to identify and classify objects or features within images. Over the last three decades, the landscape of our comprehension regarding galaxy formation and evolution has undergone a revolutionary transformation, thanks to extensive blind surveys of the sky conducted with state-of-the-art telescopes. How image detection works:1. Training Data: Algorithms are trained on large datasets that contain labeled images. These datasets help the model learnpatterns and features. 2. Neural Networks: Deep learning techniques, particularly convolutional neural networks (CNNs), are commonly used for image detection. 3.Feature Extraction: The model extracts features from the input image, such as edges, textures, and shapes. 4. Classification: Based on the learned features, the algorithm classifies the image into predefined categories.

Through a collection of labeled training images, a machine learning model undergoes a "training" process to learn and recognize a predefined set of target classes, representing items to be detected in images. In the initial stages, computer vision algorithms were fed raw pixel data. Traditional computer vision models introduced additional elements derived from pixel data, such as color histograms, textures, and shapes, to enable more flexible object modeling. However, this approach presentedchallenges, as feature engineering became cumbersome due to the substantial number of inputs requiring adjustment. Constructing robust models proved challenging due to the intricate fine-tuning of characteristics, leading to a degradation in accuracy.

The primary goal of object detection is to pinpoint points of interest, delineate these objects with rectangular bounding boxes, and as certain the specific category to which each object belongs. Examples of applications for object detection encompass identifying humans for self-driving cars, evaluating agricultural crops, and even engaging in real-time ball tracking for sports. Numerous studies and algorithms have been devised over the years to fulfill this objective. In the vast expanse of the universe, galaxies stand as celestial marvels, each uniquely shaped and composed. Understanding their diverse forms is a complex task, one that has been significantly transformed by the advent of advanced technologies in astronomy and computational sciences.

In recent times, the advancement of automated computational tools and algorithms has enabled the automated classification of galaxies and facilitated morphological analysis. Various machine learning techniques have been devised specifically for automating the morphological classification of galaxies.

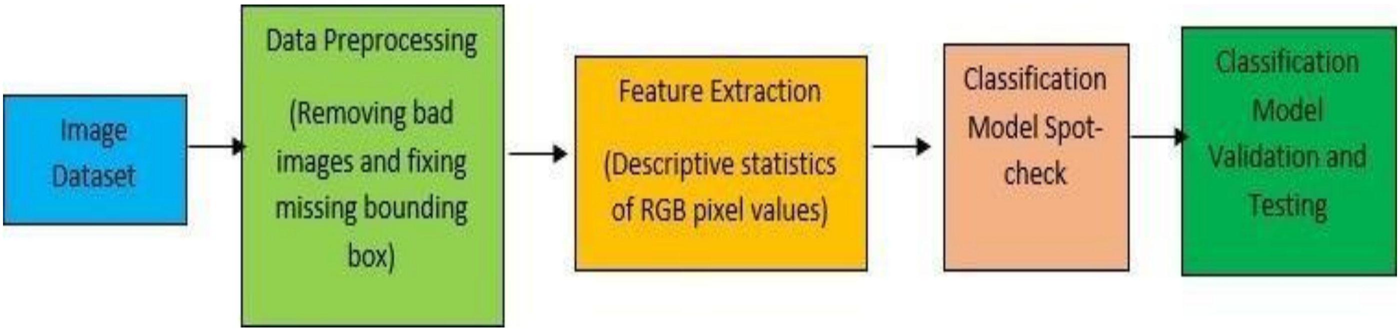

Figure 1.

II. LITERATURE SURVEY

Mohamed Abd El Aziz, I. M. Selim & Shengwu Xiong the paper presents a fresh methodology for the automated identification of galaxy morphology within datasets, employing an image- retrieval approach. This paper introduces an innovative approach for classifying galaxy types within images and simultaneously retrieving the most similar images to a given query. The proposed method comprises two key stages. Initially, a feature extraction process is conducted, leveraging shape, color, and texture descriptors. Subsequently, a binary sine cosine algorithm is employed to select the most pertinent features. In the second stage, the similarity between the features extracted from the queried galaxy image and those of other galaxy images is computed. The experimentation phase utilizes the EFIGI catalogue, housing approximately 5000 galaxy images exhibiting diverse types such as edge-on spiral, spiral, elliptical, and irregular. This approach aims to enhance notonly galaxy type detection but also the retrieval of images sharing similarity with the queried galaxyimage. This experiment was implemented in MATLAB and run in the Windows environment 64 bit [1].

Enrico M. DiTeodoro, J. E. G. Peek, John F. Wu constructed a model utilizing a deep learning algorithm, specifically a Convolutional Neural Network (CNN). This paper introduces a novel approach for catalog cleaning of galaxy fragments, presenting a newly developed method. The methodology involves training a Convolutional Neural Network (CNN) on three-band optical images to discern whether objects within catalogs represent authentic galaxy sources or mere fragments. The structure of the paper is as follows initial section outlines the utilized data and the machine-learning algorithm for galaxy fragment identification. Subsequently, the CNN is applied to some of the most comprehensive photometric catalogs available, with findings discussed in a dedicated section. The paper concludes with a summary and conclusions in the final section. To identify galaxy fragments from multiband images, a machine-learning algorithm, specifically a CNN, was adopted. Notably, individual star-formation regions within a galactic disk exhibit distinct morphological features compared to the entirety of a galaxy. The CNN proves ideal for recognizing and classifying objects based on their appearance. Functioning as a specialized neural network, the CNN employs convolutional layers to extract weighted features from input images, particularly suited for addressing the binary classification problem at hand-determining whether a target is a genuine galaxy or merely a fragment. This experiment was implemented using Python on Windows, Linux and macOS platform [2].

R.E. Gonzalez, R.P. Munoz, and C.A. Hernandez implemented model which is built upon convolutional neural networks and deep learning techniques. These approaches consistently yield superior results compared to methods relying on manual feature engineering and support vector machines (SVMs), especially in scenarios involving large training datasets. The efficacy of training results is closely tied to the conversion process employed in transforming raw FITS datafor each band into a 3- channel color image. To address this, we advocate the incorporation of data augmentation during training, leveraging five distinct conversion methods. This approach significantly enhances the overall accuracy of galaxy detection and classification, particularly when dealing with images generated from various instruments, bands, and data reduction procedures. The training of detection and classification methods is executed through the deep learning framework DARKNET and the real-time object detection system YOLO. These methodologies are implemented in the C language and CUDA platform, harnessing the computational power of graphical processing units (GPUs) for accelerated processing. This experiment was implemented on CUDA platform using C language [3].

Babatunde Keshinro implemented the model by training a machine learning model on a curated setof labeled training images. This training process equips the model to recognize a predefined set of target classes, representing the items to be detected in images. The demanding task of objectdetection involves predicting both the location and the type of objects within an image. It is crucial not only to recognize distinct images but also to estimate the concepts and positions of objects within the image, ensuring a comprehensive understanding of the overall scene. Object segmentation, in addition to object identification, involves creating a mask for the image thatdelineates the pixels constituting the object. In the early stages of computer vision, algorithms were fed with raw pixel data. Traditional computer vision models incorporated additional elements derived from pixel data, such as color histograms, textures, and shapes, to model objects in a more adaptable manner. However, this method posed challenges as feature engineering became cumbersome due to the multitude of inputs requiring adjustment. The intricate process of fine- tuning characteristics made building robust models difficult, and accuracy suffered as a consequence This experiment was implemented on Windows, Linux and macOS platform using Python [4].

Mohamed Eassa, Ibrahim Mohamed Selim, Walid Dabour, and Passent Elkafrawy utilized a nonnegative matrix factorization algorithm to create an automated supervised machine learning system designed for the classification of galaxies. The effectiveness of this technique wasevaluated using two distinct datasets comprising galaxy images. In the case of the smaller datasetcontaining 110 images, the accuracy achieved was approximately 93%. For the larger dataset consisting of 700 images, the technique demonstrated an accuracy of around 92% This experiment was implemented on Windows, Linux and macOS platform using Python [5].

John Jenkinson, Artyom Grigoryan, Mehdi Hajinoroozi, Raquel Diaz Hernandez, Hayde Peregrina Barreto, Ariel Ortiz Esquivel, Leopoldo Altamirano, and Vahram Chavushyan enacted a model that applies the previously developed theory to autonomously classify galaxies in astronomical images. The process of astronomical data collection initiates with the imaging of the sky through a camera attached to a telescope. The output from the telescope is recorded on a photographic plate, which is subsequently digitized using a high-resolution scanner operating at 6400 dpi and 16 bits per pixel. A geometric calibration is applied to rectify deformations in the photographic emulsion and correct scanning errors. The resulting digital image is then converted into a Flexible Image Transport System (FITS) format for further processing This experiment was implemented on Windows, Linux and macOS platform using Python [6].

Alexander Khalil Arwadi, Aurelio Noca, and Louis Jaugey constructed a model by leveraging the SKA SDC1 [1] dataset. This dataset was utilized to train a machine learning model designed for the detection of galaxies within these images. To achieve this objective, we introduce the pre-processing techniques employed for data cleansing. Subsequently, we delve into the intricacies of the model andthe adopted machine learning approach. The algorithm's performance is then rigorously assessed,accompanied by a comprehensive analysis of the results. Specifically, we opt to exclusively utilize images with an integration time of 1000 hours due to their lower noise levels compared to others. It's worth noting that within the training set files, galaxies are not uniformly distributed across the entirety of the images but rather confined to approximately 1% of the simulated field. This experiment was implemented on Windows and Linux platform using Python [7].

Joseph H. Murrugarra LL and Nina S. T. Hirata have shared initial findings on the classification of galaxy images through the utilization of convolutional neural networks. An accuracy in classification just above 90% was attained. To further advance this study and enhance the outcomes,our strategy involves addressing various aspects. This includes incorporating information from the 5- bands, implementing standard data augmentation techniques such as rotations, small translations, and image reflections. Additionally, we plan to combine expert-designed features with CNN features, substitute the connected component extraction method with one based on the Petrosian radius (readily available in catalogs), and expand the training set for more comprehensive learning. This experiment was implemented on Windows platform using Python [8].

Diganta Misra, Sparsha Mishra and Bhargav Appasani have published that Image processing plays a crucial role in the comprehension, examination, and interpretation of astronomical images. Itencompasses various techniques, starting with image smoothening, noise removal, edge detection, and contour mapping, up to object segmentation. When combined with signal processing, these digital image processing tools form a potent arsenal for astronomers to employ in the analysis of astronomical data. In the context of galaxy imaging, telescopes frequently capture images that feature galaxies alongside clusters of stars and other celestial objects. The initial step in the analysis process is accurately identifying the galaxy within the image proves highly beneficial in such scenarios. This technique involves identifying regional maxima and minima within the image, facilitating effective segmentation. The utilization of extrema analysis becomes particularly significant when distinguishing galaxies from surrounding celestial elements captured by telescopes. This method aids astronomers in isolating the galaxy of interest, laying the groundwork for subsequent in-depth analysis and exploration of the identified astronomical subject. This experiment was implemented using Python MATLAB and C/C++. Linux is frequently preferred dueto its stability, flexibility, and open-source characteristics [9].

Manuel Jimenez, Mercedes Torres Torres, Robert John, and Isaac Triguero provided a comparative analysis of the performance of two strategies for the automated classification of galaxy images: one involves classifying a feature set derived from the image, while the other employs convolutional neural networks (CNNs). The investigation delves into the influence of image size and the presence or absence of color channels on classification results. It also explores the impact oftwo distinct feature selection methods. The experiments are conducted using two distinct samples from the Galaxy Zoo 1 image dataset. Furthermore, the scalability of both approaches to larger datasets is examined, along with the assessment of the influence of amateur and expert classifications on classification accuracy. This experiment was implemented using both Windows and Linux using Python [10]

Lior Shamir and John Wallin have employed computational tools to automatically identifypeculiar galaxy pairs. Initially, in SDSS DR7, we identified approximately 400,000 galaxy images with a magnitude <18 that exhibited more than one point spread function. Subsequently, a machine learning algorithm was applied, identifying around 26,000 galaxy images displaying morphology akin to galaxy mergers. This dataset underwent mining using a novelty detection algorithm, resulting in a curated list of the 500 most peculiar galaxies, as determined quantitatively by the algorithm. Upon manual examination, it was observed that while not all galaxy pairs in the list werenecessarily peculiar, the approach successfully identified numerous unusual galaxy pairs. This paperdetails the protocol and computational tools employed in detecting peculiar mergers, providing illustrative examples of the identified peculiar galaxy pairs. Automatically detecting peculiar galaxies poses a complex challenge, and it is anticipated that algorithms designed for this purpose will exhibit a degree of inherent noise. Owing to this noise, a considerable number of the 500 galaxy pairs identified by the algorithm were found to lack peculiarity, with some additionally containing artifacts that had not been filtered out in the preceding stages. This experiment was implemented using Linux using Python, C/C++and Java [11].

Venkat Margapuri, Basant Thapa, and Lior Shamir used modern digital sky surveys leverage robotic telescopes to amass extensive multi-petabyte astronomical databases. Within these databases, containing billions of galaxies, the majority are classified as "regular" galaxies with known types. However, a fraction comprises rare "peculiar" galaxies that remain unidentified.These enigmatic galaxies hold significant scientific value, but their detection is challenging due to the vastness of astronomical databases. Automation becomes crucial for their identification. Given that novelty galaxies are inherently unknown, traditional machine learning models reliant on training data are ineffective. This paper introduces an unsupervised machine learning approach for the automatic detection of novelty galaxies in large databases. The method relies on a comprehensive set of numerical image content descriptors, weighted by their entropy. The farthest neighbors are ranked-ordered to address the presence of self-similar peculiar galaxies, a phenomenon expected in the vast datasets encountered in astronomy. This experiment was implemented using Windows and Linux using Python and C/C++ [12].

F. Tarsitano, C. Bruderer, K. Schawinski, and W. G. Hartley investigated the feasibility of applying machine learning techniques originally designed for 1D problems to the challenge of galaxy image classification. Traditional image classification algorithms often involve resource- intensive processes such as point spread function deconvolution, and the training and application of complex Convolutional Neural Networks with thousands or even millions of parameters. Their approach involves feature extraction from galaxy images by analyzing the elliptical isophotes in their light distribution and organizing this information into a sequence. These sequences exhibit distinctive features that enable a direct differentiation between galaxy types. Subsequently, they employ machine learning algorithms designed using the Modulo AutoML platform to train and classify these sequences. To illustrate the effectiveness of our method, we utilize the second public release of the Dark Energy Survey. Our results demonstrate a successful distinction between early- type and late-type galaxies, particularly for images with a signal-to-noise ratio exceeding 300. This leads to an accuracy of 86 percent for early-type galaxies and 93 percent for late-type galaxies, aligning with the performance of many contemporary automated image classification approaches. Notably, our novel method's data dimensionality reduction results in a substantial decrease in the computational cost of classification This experiment was implemented on Windows using Python [13].

Xuheng Ding, Simon Birrer, Tommaso Treu, and John D. Silverman provides capabilities that includes the decomposition of stellar components, as evidenced in published studies focusing on inactive galaxies and quasar host galaxies observed by instruments like the Hubble Space Telescopeand Subaru's Hyper Suprime-Cam. Galight is an open-source Python package designed for two- dimensional model fitting of optical and near-infrared images, enabling the characterization of the light distribution in galaxies. This versatile package accommodates components such as a disk, bulge, bar, and quasar in its model fitting process.Utilizing the powerful image modeling capabilities of lenstronomy, galight stands out by providinga redesigned user interface tailored for the efficient analysis of extensive datasets of extragalactic sources. The package prioritizes user-friendliness and incorporates automatic features. These include determining the cutout size for the modeling frame, locating PSF-stars within the field-of-view, estimating the noise map of the data, and identifying all objects to set initial models and associated parameters for simultaneous fitting. This experiment was implemented on Linux[14].

Jose A. de Diego, Jakub Nadolny, Angel Bongiovanni, Jordi Cepa and Mirjana Povic have accurately categorized the multitude of galaxies observed in contemporary deep surveys holds paramount importance for unraveling the mysteries of the universe and its evolutionary processes. In this context, we present our application of machine learning techniques to distinguish between early- and late-type galaxies within the OTELO and COSMOS databases, leveraging both optical and infrared photometric data. Our findings underscore the effectiveness of employing deep neural networks as a robust approach for extracting valuable insights from the extensive cataloged information. The utilization of machine learning algorithms has become increasingly prevalent inthe classification of vast astronomical databases. Notably, Linear Discriminant Analysis (LDA) emerges as a frequently employed method for classification in the realms of statistics, pattern recognition, and machine learning. LDA classifiers seek to identify optimal linear boundaries that effectively segregate the dataset, contributing to enhanced understanding and interpretation of the celestial landscape This experiment was implemented on Windows using Python [15].

Mehmet Alpaslan, Simon Driver, Aaron S. G. Robotham, Danail Obreschkow, Ellen Andrae, Michelle Cluver, Lee S. Kelvin, Rebecca Lange, Matt Owers, Edward N. Taylor, Stephen K. Andrews, Steven Bamford, and Joss Bland-Hawthorn investigated the trends in galaxy properties concerning Mpc-scale structures using datasets on environment and large-scale structure derived from the Galaxy and Mass Assembly (GAMA) survey. Our findings reveal that galaxies not situatedin groups or pairs share similar characteristics irrespective of their broader environmental context. In our sample controlled for mass, we do not observe a substantial dependency of Sersic index or galaxy luminosity on halo mass. However, a robust correlation is evident with galaxy color. When repeating our analysis for galaxies without mass control, we introduce and accentuate trends in the properties of galaxies within pairs, groups, and large-scale structures. This underscores that stellar mass emerges as the primary predictor of the examined galaxy properties, emphasizing its significance over environmental classifications This experiment was implemented on Linux and Windows using Python and R language [16].

P. H. Barchia, R. R. de Carvalho, R. R. Rosa, R. A. Sautter, M. Soares-Santos, B.A.D. Marques,E. Clua, T. S. Goncalves, C. de Sa-Freitas, and T. C. Moura researched on a distinctive approach compared to previous methods. Instead of relying on questions and answers extracted from Galaxy Zoo 2, we directly utilize the classifications and associated images. Additionally, we addressaspects overlooked in prior investigations of morphological parameters, specifically examining the threshold dependence in segmented image usage. Our study delves into the repercussions of this dependence on the parameters crucial to the TML approach. While it is well-established that Deep Learning tends to excel in perception tasks, including galaxy morphology, compared to machine learning models trained on hand-engineered features, this aspect is relatively nascent within the realm of galaxy morphology studies. Notably, a direct comparison between these two approaches has yet to be presented concurrently in the existing literature. Additionally, it is acknowledged that deep learning methods demand substantial datasets andextensive computational resources for optimal efficacy. The intricacies of tuning and managing deep learning models, coupled with potentially prolonged prediction times due to complexity, are recognized challenges. Despite these considerations, the traditional machine learning approachremains pertinent. This experiment was implemented on Windows using Python and R language [17].

H. Domínguez Sanchez, M. Huertas-Company, M. Bernardi, D. Tuccillo, and J. L. Fischer introduced a morphological catalog encompassing approximately 670,000 galaxies from the Sloan Digital Sky Survey, delineated through two perspectives: T-type, aligned with the Hubble sequence,and the Galaxy Zoo 2 (GZ2) classification scheme. Leveraging the amalgamation of precise visual classification datasets with machine learning techniques, we present the most extensive and accuratemorphological catalog to date. Employing Convolutional Neural Networks (CNNs) in deep learningalgorithms, we utilize color images from visual classification catalogues, namely GZ2 and Nair & Abraham, to train the CNNs and derive T-types. Additionally, we address a series of GZ2-type questions related to disc/features, edge-on galaxies, bar signature, bulge prominence, roundness, and mergers. Our models also yield a probability metric facilitating the segregation of pure elliptical(E) from S0 galaxies, especially in cases where the T-type model may not be as efficient. In comparison to previous models trained with support vector machines, our T-type results demonstrate reduced offset and scatter. Furthermore, their models exhibit high accuracy (>97%), precision, and recall values (>90%) when applied to a test sample mirroring the characteristics ofthe training set. This experiment was implemented on Windows using Python language [18].

R. Nevin, L. Blecha, J. Comerford, and J. Greene understood the role of merging galaxies in galaxy evolution is crucial, yet accurate identification of galaxy mergers poses a challenge. In this study, we utilize GADGET-3 hydrodynamical simulations of merging galaxies, incorporating the dust radiative transfer code SUNRISE to generate a diverse set of merging galaxy scenarios encompassing gas-poor and gas-rich mergers, as well as various mass ratios (minor and major). Adapting simulated images to SDSS imaging survey specifications, we introduce a merging galaxy classification scheme based on this imaging data. Leveraging the strengths of seven distinct imagingpredictors (Gini, M20, concentration, asymmetry, clumpiness, Sersic index, and shape asymmetry), we amalgamate them into a unified classifier employing Linear Discriminant Analysis. This integrated approach surpasses individual imaging predictors in terms of accuracy, precision, and thetimescale for observing mergers (>2 Gyr for all simulated mergers). Our findings indicate that the classification strongly depends on the mass ratio and exhibits a weaker dependence on the gas fraction of the simulated mergers. Asymmetry plays a more crucial role in major mergers, while concentration is more significant for minor mergers. This is attributed to the comparatively disrupted morphology of major mergers and the more gradual growth of stellar bulges during minor mergers. Given the substantial impact of mass ratio on the classification, we formulate distinct classification approaches for minor and major mergers applicable to SDSS imaging or adaptable forother imaging surveys. This experiment was implemented on Linux using Python and R language [19].

Caixia Zheng, Jesus Pulido, Paul Thorman, and Bernd Hamann presents an enhanced approach for identifying celestial objects (galaxies and stars) in astronomical images. The method begins witha global detection scheme, followed by a more precise refinement achieved by segmenting the entire image into irregularly sized sub-regions using the watershed segmentation method. Within each sub-region, a sophisticated detection procedure is applied, incorporating adaptive noise reduction and a layered strategy to identify bright and faint objects, respectively. Subsequently, a multi-threshold technique is employed to separate merged objects. When tested on simulated data, this method demonstrates the capability to detect more genuine objects compared to SEXTRACTOR, achieving a true detection rate of 91 percent compared to 83 percent with SEXTRACTOR, and exhibiting an improved ability to detect extremely faint objects, up to 2 magnitudes fainter than SEXTRACTOR on similar data. The effectiveness of our method is further validated through its application to real observational image datasets. This experiment was implemented on any operating system like Linux Windows using Python and R language [20].

III. METHODOLOGY

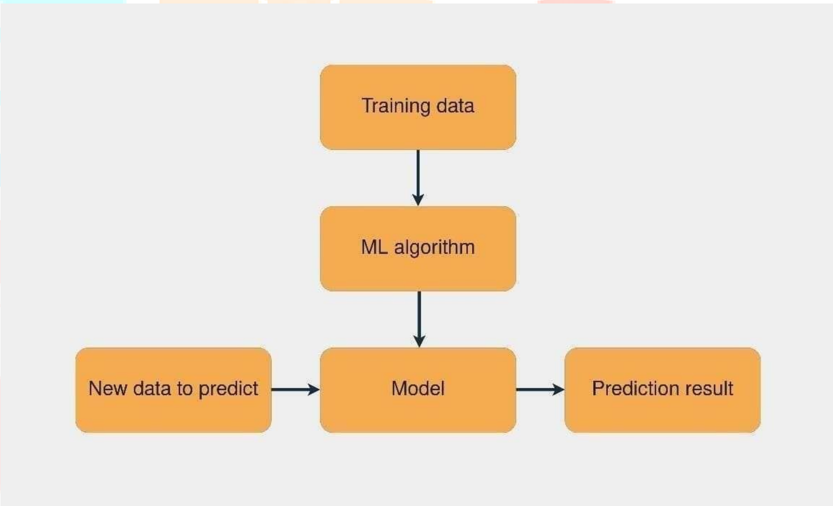

Figure 2.

The initial step involves "validating" the model using a limited dataset. Following this, the model undergoes a "training" phase with a more extensive dataset, capturing its behavior across diverse situations and circumstances. After successful training, the model is ready for deployment. When presented with new data, it leverages its learned behavior to "predict" outcomes based on theinput data. This description is simplified, recognizing that real-world applications involve various variations and nuances. As depicted in the straightforward process diagram above, data assumes a pivotal role throughout every phase of development and training. The data provided to the model not only influences the content of the model's output but also determines its effectiveness. In essence, the model's performance mirrors the quality of the input data.

In the depicted flowchart (Fig 2), three essential steps crucial for executing the model are distinctly classified as follows:

1. **Identification:** In the process of galaxy detection, the identification stage involves the examination of celestial data to pinpoint and characterize galaxies within a given astronomical datasets. This crucial step employs advanced algorithms and pattern recognition techniques to differentiate galaxies from other celestial

objects, ensuring accurate and reliable identification. Through this process, distinct features, such as shape, luminosity, and spectral characteristics, are analyzed to provide a comprehensive understanding of the galaxies present in the observed regions of the universe.

2. **Training:** During the training phase of galaxy detection, a complex procedure takes place as machine learning models are introduced to a varied array of astronomical data. This training process includes the refinement of models through advanced algorithms, allowing them toaccurately recognize and classify galaxies. The models develop the capacity to identify patterns,features, and subtleties within the data, enabling them to generalize and identify galaxies in previously unseen observations. Through this iterative learning process, the model's performance is honed, bolstering its ability to effectively detect and categorize galaxies in diverse celestial and scrapes.

3. **Validation:** During the validation phase of galaxy detection, extensively trained machine learning models undergo rigorous testing and evaluation to confirm their effectiveness beyond the confines of the training data. This entails exposing the models to a distinct dataset that they have not encountered previously. The objective is to assess their capability to generalize and precisely identify galaxies within varied and previously unexplored astronomical contexts. The validation process is instrumental in identifying potential over fitting or under fitting concerns and plays a crucial role in refining the model parameters for optimal performance. It serves as a pivotal step in testing the model's dependability and efficiency in practical applications for detecting galaxies in the real-world astronomical landscape.

These fundamental steps pave the way for the straightforward identification of galaxies in spatial arrangements during the complex process of detection.

Dataset A dataset is like a structured collection of information. In tech and data science, it's a bunch of data points or observations all about a specific thing. These collections can have numbers,categories, or words. People use datasets to do lots of things, like teaching and testing computers, doing math and stats, and finding out interesting stuff from the data. Datasets can be big or small, depending on how much info is needed for a certain job.

The dataset utilized in our model is GalaxyZoo (GZ):
https://astronn.readthedocs.io/en/latest/galaxy10.html

IV. CONCLUSION

Implementation could have been on a broader perspective scale including different classes ofgalaxies. Rather than only predicting the name, shape and contrast of the galactical images, entire sequence of information could have been retrieved. The location of galaxies could also have been computed. The information regarding the users of the application could have been stored in database for maintaining purpose.

IV. REFERENCES

[1]. Mohamed Abd El Aziz, I. M. Selim & Shengwu Xiong, "Automatic Detection of Galaxy Type from Datasets of Galaxies Image Based on Image Retrieval Approach", Scientific Reports, 30th June 2017.

[2]. Enrico M. DiTeodoro, J. E. G. Peek, John F. Wu, "Identification of Galaxy Shreds in Large Photometric Catalogs Using Convolutional Neural Networks", The Astronomical Journal, 22nd February2023.

[3]. R.E. Gonzalez, R.P. Munoz, and C.A. Hernandez, "Galaxy detection and identification using deep learning and data augmentation", Astronomy and Computing, 15th September 2018.

[4]. Babatunde Keshinro, "Image Detection and Classification: A Machine Learning Approach", Social Science Research Network (SSRN), 13th December 2022.

[5]. Mohamed Eassa, Ibrahim Mohamed Selim, Walid Dabour, and Passent Elkafrawy, "Automated detection and classification of galaxies based on their brightness patterns", Alexandria Engineering Journal, 26th June 2018.

[6]. John Jenkinson, Artyom Grigoryan, Mehdi Hajinoroozi, Raquel Diaz Hernandez, Hayde Peregrina Barreto, Ariel Ortiz Esquivel, Leopoldo Altamirano, and Vahram Chavushyan, "Machine Learning and Image Processing in Astronomy with Sparse Data Sets", ResearchGate, 25th January 2015.

[7]. Alexander Khalil Arwadi, Aurelio Noca, and Louis Jaugey, "Galaxy Detection Machine Learning Project", Ecole Polytechnique Federale de Lausanne (EPFL), May 2021.

[8]. Joseph H. Murrugarra LL and Nina S. T. Hirata, "Galaxy Image Classification", SIBGRAPI, 2017.

[9]. Diganta Misra, Sparsha Mishra and Bhargav Appasani, "Advanced Image Processing for Astronomical Images", Arxiv,23rd December 2018.

[10]. Manuel Jiménez, Mercedes Torres Torres, Robert John, and Isaac Triguero, "Galaxy Image Classification Based on Citizen Science Data: A Comparative Study", Institute of Electrical and Electronics Engineers (IEEE), 5th March 2020.

[11]. Lior Shamir and John Wallin, "Automatic detection and quantitative assessment of peculiar galaxy pairs in Sloan Digital Sky Survey", Royal Astronomical Society, 9th August 2014.

[12]. Venkat Margapuri, Basant Thapa, and Lior Shamir, "Detection of unknown galaxy types in large databases of galaxy images", Research Gate, Published Online on 10th August 2020.

[13]. F. Tarsitano, C. Bruderer, K. Schawinski, and W. G. Hartley, "Image feature extraction and galaxy classification: a novel and efficient approach with automated machine learning", Royal Astronomical Society, 5th May 2021.

[14]. Xuheng Ding, Simon Birrer, Tommaso Treu, and John D. Silverman, "Galaxy shapes of Light (GaLight): a 2D modeling of galaxy images", Arxiv,16th November 2021.

[15]. Jose A. de Diego, Jakub Nadolny, Angel Bongiovanni, Jordi Cepa and Mirjana Povic, "Galaxy classification: deep learning on the OTELO and COSMOS databases", EDP Sciences, 25th April 2020.

[16]. Mehmet Alpaslan, Simon Driver, Aaron S. G. Robotham, Danail Obreschkow, Ellen Andrae, Michelle Cluver, Lee S. Kelvin, Rebecca Lange, Matt Owers, Edward N. Taylor, Stephen K. Andrews, Steven Bamford, and Joss Bland-Hawthorn, "Galaxy and Mass Assembly (GAMA): trends in galaxy colors, morphology, and stellar populations with large-scale structure, group, and pair environments", Royal Astronomical Society, 4th March 2015.

[17]. P. H. Barchia, R. R. de Carvalho, R. R. Rosa, R. A. Sautter, M. Soares-Santos, B.A.D. Marques, E. Clua, T. S. Gonçalves, C. de Sa-Freitas, and T. C. Moura, "Machine and Deep Learning Applied to Galaxy Morphology - A Comparative Study", Arxiv,1st November 2019.

[18]. H. Domínguez Sanchez, M. Huertas-Company, M. Bernardi, D. Tuccillo, and J. L. Fischer,"Improving galaxy morphologies for SDSS with Deep Learning", Royal Astronomical Society, 7th September 2017.

[19]. R. Nevin, L. Blecha, J. Comerford, and J. Greene, "Accurate Identification of Galaxy Mergers with Imaging", The Astrophysical Journal, 12th February 2019.

[20]. Caixia Zheng, Jesus Pulido, Paul Thorman, and Bernd Hamann, "An improved method for object detection in astronomical images", Royal Astronomical Society, 14th January 2015.