1. INTRODUCTION

Social media technology has revolutionized the landscape of both personal and professional communication, and social media platforms are now an almost vital part of most modern human personal lives [1]. Twitter is one of the social media platforms that is widely used by modern people. Twitter offers a medium for all individuals to express themselves freely and opens a place to hear various kinds of expressions and voices that are spread by many people. Ease of access must be followed by online responsibility and the ability to understand existing regulations to create a clean social media environment. These problems are difficult to handle because it is difficult to manage the activities carried out by the user. It is the responsibility of each individual. We must create a safer place for social media environments and avoid the spread of hate speech. A challenging problem that arises in this domain is crucial and requires considerable efforts to improve online responsibility and balance freedom of expression. A highly accurate hate speech detection system should be implemented as soon as possible.

To overcome this problem, some approaches have been made to detect hate speech [2]–[8]. A recent idea in the development of a hate speech detection system is to utilize hybrid deep learning and feature expansion to reduce word mismatches in datasets. However, in previous research, they still used conventional machine learning and solo deep learning in hate speech detection [9], [10]. As far as we know, there is still less research on hate speech detection that utilizes hybrid deep learning and feature expansion. Hybrid deep learning itself is a combination of two or more different deep learning methods and is very useful for training large amounts of data. Deep learning aims to mimic the human brain's ability to create and maintain representations of its environment that predict possible outcomes based on user data, allowing machines to display behavior learned from experience rather than human interaction [11]. Semantic vectors contain many linguistic features that may have features in common with one another. Feature expansion is one of the new methods to reduce vocabulary mismatches that happen in the semantic vector by identifying missing words and replacing them with semantically similar words [12].

Research on hate speech carried out by Melton et al. [13] proposed a combination of deep learning approaches with three different datasets. One of the exciting parts of their study is implementing an ensemble that combines CNN, RNN, combination FC. What they don't realize is the use of pre-trained models, namely CommonCrawl and Wiki, in extraction using FastText or GloVe. The pre-trained model used is not specific for hate speech detection and, of course, consists of many languages, and the study was overly optimistic. Attention mechanisms and deep learning were used in research [14]. The author in this study uses hybrid ensemble deep learning with CNN and Bi-GRU. This is unique because they built a binary classification voting system. The author stated that the voting system and the addition of an attention mechanism to the hybrid layer had a major effect on increasing the accuracy of the model. The attention mechanism certainly has drawbacks in terms of computer complexity and calculations, but these deficiencies are covered by its many advantages, such as making it easier for the model to recognize slugs and slang terms in hate speech. Hybrid deep learning approaches were used in research carried out by Elzayady et al. [15]. Their research developed an automated method based on personality literature to identify Arabic hate speech, and they state that their research is the first in this regard.

Several studies have also been conducted on feature expansion [16]–[19]. In one of the studies on hate speech detection [18], GloVe was utilized for feature expansion. The classification still uses machine learning, namely, Logistic Regression (LR), Random Forest (RF), and Artificial Neural Network (ANN). The result shows that feature expansion with a combination of TF-IDF and corpus tweets built on GloVe provides an average accuracy value of 88.59%. Feature expansions were used by Ghozali et al. [20] for the detection of hate speech in Indonesian languages. Their concept of feature expansion is to find synonymous words and add all the synonymous words that they find to the features. The lack of concepts carried out by this author has an impact on computer calculations and the complexity of the algorithms. The addition of another feature causes the current features to become more numerous and uncontrollable, which places a burden on the model. It is important to exercise proper feature selection and consider the trade-off between model complexity and performance. Feature expansion is a challenge and one way to select features. The selection of the correct algorithms and techniques is very necessary for improving the concept of feature expansion.

This study proposes an approach to detect hate speech using a deep learning approach with feature expansion to leverage linguistic richness. Bidirectional Gated Recurrent Units (Bi-GRU) and Convolutional Neural Networks (CNN) are two deep learning methods used in this study. In general, we use deep learning and hybrid approaches as in previous studies, but we added a feature expansion that has a different concept from [20] and a comparison between SDL and HDL. Our concept is that feature expansion is carried out in a semantic vector to replace missing words with semantically similar words using the help of a self-made corpus using FastText. In summary, the contribution of this paper is as follows: i) comparison between solo deep learning and hybrid deep learning in terms of understanding hate speech more comprehensively; ii) presentation of our feature expansion algorithm concept, which is performed in semantic vector to detect hate

speech to minimize computer calculations and the complexity of algorithms; iii) successfully implementing a good model without overfitting. This approach would represent a breakthrough in hate speech detection. To achieve our research target, this study takes several steps. Building a baseline, or basic model, is the first step in achieving our target. The baseline model is then used as a benchmark model for the next step, in which there are steps for feature expansion and the application of various types of n-grams with TF-IDF and dropout.

The subsequent section of this study is Section II, which will delve into the methodology employed in this research. Section III will encompass the findings and discussion of this study. Lastly, Section IV comprises the conclusion, recommendations, and prospects for future research endeavors aimed at enhancing the accuracy of hate speech detection.