(Improving Cyberbullying Detection with User Context) this paper shows that taking user context into account improves the detection of cyberbullying. 4626 comments from 3858 distinct users. The comments were manually labelled as bullying (9.7%) and non-bullying (inter-annotator agreement 93%). They applied SVM classifier and were able to reach results of up to 78% on precision and 55% on recall.

(Cyberbullying Detection with a Pronunciation Based Convolutional Neural Network) this paper uses novel pronunciation based convolutional neural network (PCNN) thereby alleviating the problem of noise and bullying data sparsity to overcome class imbalance. 1313 messages from twitter, 13,000 messages from formspring.me. Accuracy of the twitter dataset wasn’t calculated due to it being imbalanced. While Achieving 56% on precision, 78% recall and 96% accuracy.

(Experts and Machines against Bullies: A Hybrid Approach to Detect Cyberbullies) this paper applied 3 models for their dataset gathered YouTube comment section, Multi-Criteria Evaluation Systems (MCES), machine learning: (Naive Bayes classifier, decision tree, SVM), Hybrid approach. The MCES score 72% on accuracy, while Naive Bayes scored the highest out of the three with 66%.

(A Pattern-Based Approach for Sarcasm Detection on Twitter) this paper put most of its effort in Natural Language Processing (NLP) and SVM, their dataset is collected from twitter. While reaching 83.1% in accuracy, and 91.1% in precision.

(Unsupervised Cyber Bullying Detection in Social Networks) this paper proposed to adopt an unsupervised approach to detect cyber bully traces over social networks. They used the classifiers inconsistently over their dataset, applying SVM on FormSpring and achieving 67% on recall, applying GHSOM on YouTube and achieving 60% precision, 69% accuracy and 94% recall,  
applying Naive Bayes on Twitter and achieving 67% accuracy.

(Cyberbullying System Detection and Analysis) this paper showed promising results due to relying on heavy feature extraction information gain using tfidf, LIWC and unusual capitalization. Their data is collected from Ask.fm. All being passed to the SVM classifier and scoring 99.4% in accuracy, 69% in precision, 84.9% in recall.

(Common Sense Reasoning for Detection, Prevention, and Mitigation of Cyberbullying) this paper their aim was to detect explicit bullying language pertaining to (1) sexuality, (2) race & culture and (3) intelligence, acquiring their dataset from the YouTube comment section.  
They applied SVM and reached 66% in accuracy and Naive Bayes reached 63% in accuracy.

(Improved Cyberbullying Detection Using Gender Information) this paper used a supervised learning approach to detect cyberbullying. They constructed a Support Vector Machine classifier using WEKA. Their dataset was collected from Myspace. They achieved 43% on precision, 16% in recall and their accuracy wasn’t mentioned.

(Cyberbullying Detection using Time Series Modeling) this paper unlike previous approach that consider a fixed window of a cyber-predator’s questions within a dialogue, we exploit the whole question set and model it as a signal, whose magnitude depends on the degree of bullying content, their pre-processing: Stop-words removal, tokenization, stemming and Part-Of-Speech tagging. Applying SVM achieved a 49.8% result on accuracy.

(Machine Learning Approach for Detection of Cyber-Aggressive Comments by Peers on Social Media Network) In this paper we have devised methods to detect cyberbullying using supervised learning techniques. We present two new hypotheses for feature extraction to detect offensive comments directed towards peers which are perceived more negatively and result in cyberbullying. Their Feature Extraction: N-gram model, Counting: Count the number of times each of these tokens occurs in each of the text strings, TF-IDF Score. They used SVM and reached results of 70.29% on precision and 58.29% on recall. Logistic Regression scored 64.4% on precision and 61.47% on recall.

(A Multilingual System for Cyberbullying Detection: Arabic Content Detection using Machine Learning) None till now covered cyberbullying in Arabic language and this paper wants to tackle this issue. Features used: Unigrams, TF-IDF, Lexicon, Bigrams. Machine Learning: Naive Bayes Scored 90.8514 %, Support Vector Machine (SVM) Scored 94.1%.