The paper I chose to discuss here is titled **Mastering the game of Go with deep neural networks and tree search**, which was written by David Silver, etc. and published in ***Nature*** on 28 January 2016. The program described in this paper is the very famous AlphaGo, which is almost known by everyone. This Go program plays at the level of the strongest human players, achieving one of the artificial intelligence's grand challenges.

Before AlphaGo, the strongest Go programs are based on Monte Carlo tree search, enhanced by policies that are trained to predict human expert moves. These prior works have been limited to shallow policies or value functions based on a linear combination of input features.

AlphaGo program uses a new approach of computing Go that has the combination of deep neural network and tree search to select moves. Deep convolutional neural networks achieved unprecedented performance in visual domains but not in game playing yet before AlphaGo. In AlphaGo, the board position is passed as a 19 X 19 image and uses convolutional layers to construct a representation of the position, then trains the neural network with a novel combination of supervised learning from human expert games, and reinforcement learning from games of self-play.

The pipeline of AlphaGo begins by training a supervised learning (SL) policy network $p\sigma$ directly from expert human moves. This provides fast, efficient learning updates with immediate feedback and high-quality gradients. Similar to prior work, it also trains a fast policy $p\pi$ that can rapidly sample actions during rollouts. Next, it trains a reinforcement learning (RL) policy network $p\rho$ that improves the SL policy network by optimizing the final outcome of games of self-play. This adjusts the policy towards the correct goal of winning games, rather than maximizing predictive accuracy. Finally, we train a value network $v\theta$ that predicts the winner of games played by the RL policy network against itself. AlphaGo efficiently combines the policy and value networks with MCTS. This way effectively reduces the depth and breadth of the search tree. Without any look ahead search, the neural networks play Go at the level of state-of-the-art Monte Carlo tree search programs that simulate thousands of random games of self-play.

This paper also showed the evaluate results of AlphaGo. The results of the tournament suggest that single-machine AlphaGo is many *dan* ranks stronger than any previous Go program, at winning rate of 99.8% against other Go programs. Most importantly, the distributed version of AlphaGo is evaluated by against Fan Hui, a professional 2 dan, and the winner of the 2013, 2014 and 2015 European Go championship. AlphaGo won the match 5 games to 0. This is the first time that a computer Go program has defeated a human professional player, without handicap, in the full game of Go- a feat that was previously believed to be at least a decade away.