

1.1 Latar Belakang

Tuberculosis (TB) sampai saat ini masih menjadi masalah kesehatan masyarakat utama di banyak Negara yang dapat menyebabkan kematian (Leshinsky, 2016) . Selama dekade terakhir, jumlah pasien yang terinfeksi tuberculosis terus meningkat setiap tahunnya (Lange, Dheda, Chesov, Mandalakas, Udwadia, & Jr, 2019). Tuberculosis adalah infeksi bakteri di udara yang disebabkan oleh *Mycobacterium tuberculosis* (MTB) yang menyerang bagian tubuh mana pun dan sebagian besar umumnya adalah paru-paru, tanda-tanda gejala tuberculosis yaitu : batuk sering berlangsung lebih lama dari 3 minggu dengan atau tanpa produksi dahak, batuk darah, sakit dada, kehilangan nafsu makan, penurunan berat badan yang tak terduga, keringat malam, demam dan kelelahan (Agyeman & Richard, 2017). Tuberculosis ditularkan melalui udara ketika penderita batuk paru-paru (Dunn, Starke, & Revella, 2016). Diagnosis tuberculosis yang cepat dan akurat sangat penting untuk memastikan terapi dan pengobatan yang tepat (Jeong, Lee, & Jae-Joon, 2017).

Penelitian bidang bioinformatika menjadi populer saat ini sebagai solusi bagi dunia medis (Putra, Jondri, & Fitriyani, 2018), dengan perkembangan teknologi di bidang medis saat ini, mengakibatkan *machine learning* menjadi salah satu metode yang sangat diminati dimana dengan bantuan sistem ini kemungkinan akan kesalahan diagnosis yang dilakukan oleh para ahli medis dapat dihindari.(Nugraha, Shidiq, & Rahayu, 2019). Beberapa algoritma machine learning yang populer adalah K-Nearest Neighbor (KNN), support vector machine (SVM), Naive Bayes, Radial Basis Function (RBF), LDA dan Learning Vector Quantization (LVQ). Machine learning telah banyak digunakan dalam bidang medis untuk menganalisa dataset medis, (Waspada, Wibowo, & Meraz, 2017) menganalisis dataset gen kanker pada pasien dengan menggunakan algoritma *machine learning supervised* dalam mengklasifikasikan sel kanker berdasarkan ekspresi gen microRN. (Chazar & Widhiaputra, 2020) menganalisis dataset kanker payudara menggunakan algoritma support vector machine (SVM) algoritma SVM pada ML digunakan untuk mencari suatu pola data dari sekumpulan data yang dapat menghasilkan prediksi untuk menentukan sel hidup kanker payudara bersifat ganas atau jinak. (Wibowo, 2017) menganalisis dataset kanker payudara menggunakan algoritma sequential minimal optimization.

Pada Penelitian ini akan digunakan algoritma *machine learning* K-Nearest Neighbor (KNN) dalam mengklasifikasikan *tuberculosis* (TB) dikarenakan K-Nearest Neighbor (KNN) memiliki beberapa kelebihan yaitu tangguh terhadap training data yang noisy dan efektif apabila

di proses pada data latih yang besar (Musa & Alang, 2017). File citra x-ray yang digunakan untuk pelatihan maupun pengujian diambil dari http://openi.nlm.nih.gov/imgs/collections/ChinaSet_AllFiles.zip dengan total keseluruhan sampling 1000 citra penyakit *tuberculosis* (TB), terdapat beberapa penelitian terkait dengan menerapkan algoritma K-Nearest Neighbor (KNN) dan ekstraksi fitur SURF : (Handayani, 2019) Mengklasifikasikan penyakit Disk Hernia dan Spondylolisthesis pada tulang belakang. Data disusun dalam dua tugas yang berbeda namun terkait, yaitu kategori “normal” dan “abnormal” hasil penelitian menunjukkan bahwa pengujian dari pengklasifikasi K-NN sebesar 83%. Rata-rata lama waktu yang diperlukan untuk melakukan klasifikasi classifier K-NN 0,000212303 detik. (Musa & Alang, 2017) menganalisis penyakit paru-paru menggunakan algoritma K-NN pada rumah sakit aloe saboe akota gorontalo hasil prediksi yang dilakukan oleh algoritma K-NN menghasilkan akurasi yang cukup tinggi mencapai 91.90%. (Mustafa & Simpen, 2019) menggunakan algoritma K-Nearest Neighbor (KNN) untuk memprediksi pasien terkena penyakit diabetes pada puskesmas manyampa kabupaten bulukumba hasil akurasi yang diperoleh dari pengujian tersebut sebesar 68,30%. (Ariel, Atmaja, & Azizah, 2016) pada penelitiannya menggunakan ekstraksi fitur SURF dan SIFT untuk mendapatkan ciri telapak kaki seseorang dimana hasil pengujian yang telah diperoleh, maka ekstraksi ciri SURF lebih baik digunakan dibandingkan ekstraksi ciri SIFT dalam sistem pengidentifikasian individu menggunakan citra telapak kaki berdasarkan ketepatan dalam mengelompokkan citra. (S & Mathew, 2016) dalam penelitiannya membandingkan algoritma SIFT dan SURF untuk pengenalan sistem biometris iris hasil percobaan membuktikan bahwa algoritma SURF bekerja lebih cepat, efisien dan dapat diandalkan daripada SIFT.

1.2 Rumusan Masalah

Berdasarkan latar belakang sebelumnya maka rumusan masalah yang dapat diambil adalah menganalisis penyakit *tuberculosis* (TB) pada citra x-ray menggunakan ekstraksi fitur SURF dan metode klasifikasi K-Nearest Neighbor (KNN) untuk mendapatkan hasil negative dan positif penyakit *tuberculosis* (TB).

1.3 Tujuan dan Manfaat

1. Untuk mendapatkan persentase keakurasian pada metode K-Nearest Neighbor (KNN) dalam mengenali citra x-ray penyakit *tuberculosis* (TB) dengan ekstraksi fitur SURF
2. Sebagai evaluasi perbandingan terhadap berbagai pendekatan lainnya.

1.4 Batasan Masalah

1. File citra yang digunakan berformat .bmp
2. Berasal dari http://openi.nlm.nih.gov/imgs/collections/ChinaSet_AllFiles.zip
3. Menggunakan ekstraksi fitur SURF
4. Metode klasifikasi yang digunakan K-Nearest Neighbor (KNN).

1.5 Keterbaruan

(Waspada, Wibowo, & Meraz, 2017) menganalisis dataset gen kanker pada pasien dengan menggunakan algoritma *machine learning supervised* dalam mengklasifikasikan sel kanker berdasarkan ekspresi gen microRN. (Chazar & Widhiaputra, 2020) menganalisis dataset kanker payudara menggunakan algoritma support vector machine (SVM) algoritma SVM pada ML digunakan untuk mencari suatu pola data dari sekumpulan data yang dapat menghasilkan prediksi untuk menentukan sel hidup kanker payudara bersifat ganas atau jinak. (Wibowo, 2017) menganalisis dataset kanker payudara menggunakan algoritma sequential minimal optimization. (Handayani, 2019) Mengklasifikasikan penyakit Disk Hernia dan Spondylolisthesis pada tulang belakang. Data disusun dalam dua tugas yang berbeda namun terkait, yaitu kategori “normal” dan “abnormal” hasil penelitian menunjukkan bahwa pengujian dari pengklasifikasi K-NN sebesar 83%. Rata-rata lama waktu yang diperlukan untuk melakukan klasifikasi classifier K-NN 0,000212303 detik. (Musa & Alang, 2017) menganalisis penyakit paru-paru menggunakan algoritma K-NN pada rumah sakit aloe saboe akota gorontalo hasil prediksi yang dilakukan oleh algoritma K-NN menghasilkan akurasi yang cukup tinggi mencapai 91.90%. (Mustafa & Simpen, 2019) menggunakan algoritma K-Nearest Neighbor (KNN) untuk memprediksi pasien terkena penyakit diabetes pada puskesmas manyampa kabupaten bulukumba hasil akurasi yang diperoleh dari pengujian tersebut sebesar 68,30%. (Ariel, Atmaja, & Azizah, 2016) pada penelitiannya menggunakan ekstraksi fitur SURF dan SIFT untuk mendapatkan ciri telapak kaki

seseorang dimana hasil pengujian yang telah diperoleh, maka ekstraksi ciri SURF lebih baik digunakan dibandingkan ekstraksi ciri SIFT dalam sistem pengidentifikasian individu menggunakan citra telapak kaki berdasarkan ketepatan dalam mengelompokkan citra. (S & Mathew, 2016) dalam penelitiannya membandingkan algoritma SIFT dan SURF untuk pengenalan sistem biometris iris hasil percobaan membuktikan bahwa algoritma SURF bekerja lebih cepat, efisien dan dapat diandalkan daripada SIFT.