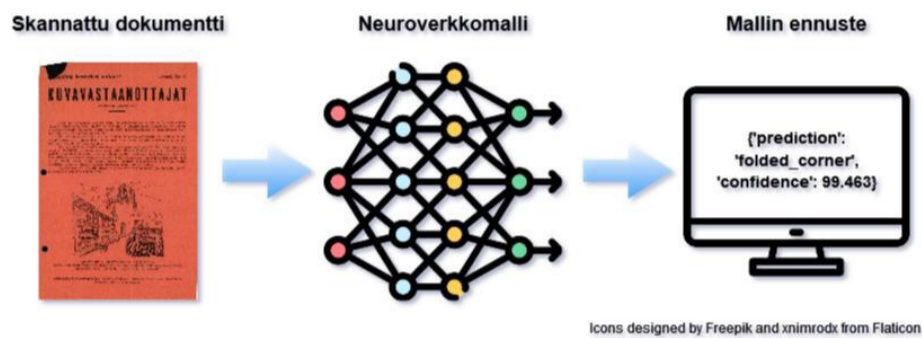


## Taittuneiden kulmien tunnistus -komponentti

### Komponentin tausta, hyöty ja käyttötarkoitus

Asiakirjojen digitointiprosessissa skannatuksi saattaa tulla asiakirjoja, joissa paperin kulma tai reuna on taittunut tai revennyt. Mikäli taitos tai repeämä vaikuttaa asiakirjan sisällön luettavuuteen, on hyvä, jos tällaiset tapaukset voidaan tunnistaa ja tarvittaessa skannata uudelleen. Taittuneiden kulmien koneellinen tunnistus auttaa näin parantamaan digitaalisen aineiston laatua ja vähentää tarvetta manuaaliseen laaduntarkistukseen.

### Mitä komponentti tekee?



Komponentti käsittelee syötteenä saamansa kuvatiedostot yksi kerrallaan ja muodostaa koulutuksen aikana oppimiensa parametrien pohjalta ennustuksen kuvan sisällöstä. Mikäli komponentti tunnistaa kuvasta taitoksen tai repeämän yli 50 % todennäköisyydellä, kuva luokitellaan virheelliseksi. Komponentti palauttaa käyttäjälle ennustetun luokan.

### Miten komponenttia käytetään?

Komponentti on saatavilla sekä osana Arkkiivi-käyttöliittymää (<http://www.arkkiivi.fi/>) että itsenäisenä, ohjelmointirajapinnan (API) tarjoavana sovelluksena DALAI-hankkeen GitHub-sivulla (<https://github.com/DALAI-project/CornerAPI>). Tarkemmat ohjeet komponentin käyttöön eri ympäristöissä löytyvät oheisilta verkkosivuilta.

### Komponentin koulutus

Komponentin koulutukseen on käytetty .jpg-muotoisia kuvatiedostoja, joiden koko on yhdenmukaistettu 224 x 224 pikseliin. Koulutuksessa käytettyjä dokumentteja on yhteensä n. 35 000 kappaletta, joista taitoksia tai repeämiä sisältäviä dokumentteja on n. 5 000. Koulutusaineisto on valikoitu niin, että komponentti tunnistaisi dokumenteista taittuneita ja revenneitä kulmia riippumatta siitä, vaikuttavatko virheet suoraan dokumentin sisällön luettavuuteen (esim. taittuneen kulman ei tarvitse peittää tekstisisältöä).

Vaikka koulutusaineistoon on pyritty kokoamaan monipuolisesti erilaisia esimerkitapauksia, aineiston määrä on rajallinen ja komponentin luokittelussa tapahtuu myös virheitä. Esimerkiksi mikäli dokumentin kulma muistuttaa värityksensä tai muotonsa takia taitosta, saattaa tämä johtaa virheelliseen luokitukseen.