Effect of galaxy mergers on star-formation rates

W. J. Pearson[1,2], L. Wang[1,2], M. Alpaslan[3], I. Baldry[4], M. Bilicki[5], M. J. I. Brown[6], M. W. Grootes[7],
B. W. Holwerda[8], T. D. Kitching[9], S. Kruk[10], and F. F. S. van der Tak[1,2]

1   SRON Netherlands Institute for Space Research, Landleven 12, 9747 AD Groningen, The Netherlands
    e-mail: `w.j.pearson@sron.nl`
2   Kapteyn Astronomical Institute, University of Groningen, Postbus 800, 9700 AV Groningen, The Netherlands
3   Center for Cosmology and Particle Physics, New York University, 726 Broadway, New York, NY 10012, USA
4   Astrophysics Research Institute, Liverpool John Moores University, Twelve Quays House, Egerton Wharf,
    Birkenhead CH41 1LD, UK
5   Center for Theoretical Physics, Polish Academy of Sciences, Al. Lotników 32/46, 02-668 Warsaw, Poland
6   School of Physics and Astronomy, Monash University, Clayton, Victoria 3800, Australia
7   Netherlands eScience Center, Science Park 140, 1098 XG Amsterdam, The Netherlands
8   Department of Physics and Astronomy, 102 Natural Science Building, University of Louisville, Louisville, KY 40292, USA
9   Mullard Space Science Laboratory, Dorking RH5 6NT, UK
10 European Space Agency, ESTEC, Keplerlaan 1, 2201 AZ Noordwijk, The Netherlands

ABSTRACT

Context. Galaxy mergers and interactions are an integral part of our basic understanding of how galaxies grow and evolve over time. However, the effect that galaxy mergers have on star-formation rates (SFRs) is contested, with observations of galaxy mergers showing reduced, enhanced, and highly enhanced star formation.
Aims. We aim to determine the effect of galaxy mergers on the SFR of galaxies using statistically large samples of galaxies, totalling over 200 000, which is over a large redshift range from 0.0 to 4.0.
Methods. We trained and used convolutional neural networks to create binary merger identifications (merger or non-merger) in the SDSS, KiDS, and CANDELS imaging surveys. We then compared the SFR, with the galaxy main sequence subtracted, of the merging and non-merging galaxies to determine what effect, if any, a galaxy merger has on SFR.
Results. We find that the SFR of merging galaxies are not significantly different from the SFR of non-merging systems. The changes in the average SFR seen in the star-forming population when a galaxy is merging are small, of the order of a factor of 1.2. However, the higher the SFR is above the galaxy main sequence, the higher the fraction is for galaxy mergers.
Conclusions. Galaxy mergers have little effect on the SFR of the majority of merging galaxies compared to the non-merging galaxies. The typical change in SFR is less than 0.1 dex in either direction. Larger changes in SFR can be seen but are less common. The increase in merger fraction as the distance above the galaxy main sequence increases demonstrates that galaxy mergers can induce starbursts.

Key words. galaxies: interactions – galaxies: evolution – galaxies: star formation – galaxies: starburst – methods: numerical

1. Introduction

Galaxy mergers and interactions form a key part of our understanding of how galaxies form and evolve over time. In cold dark matter cosmology, dark matter halos merge under hierarchical growth that results in the merger of the halos' baryonic counterparts (e.g. Conselice 2014; Somerville & Davé 2015). This interaction results in the disruption of the galaxies that lie at the centre of the dark matter halos. Tidal forces act to pull and distort the galaxies, subsequently moving stars within the galaxies from the disk to the spheroid component (e.g. Toomre & Toomre 1972; Somerville & Davé 2015, and references therein). Mergers can potentially increase the activity of an active galactic nucleus (e.g. Sanders & Mirabel 1996; Ellison et al. 2019), although more recent work suggests this may not always be the case (e.g. Darg et al. 2010a; Weigel et al. 2018).

Mergers are also thought to trigger periods of extreme star formation: starbursts. From simulations, these starbursts are believed to be a result of the tidal interactions of the galaxies that compress and shock the gas, which results in the rapid formation of stars (e.g. Barnes 2004; Kim et al. 2009; Saitoh et al. 2009). Such shock-induced star formation in mergers has also been observed (Schweizer 2009). These intense star-forming events are believed to be the cause of some of the brightest infrared objects, ultra luminous infrared galaxies (Sanders & Mirabel 1996; Niemi et al. 2012). This connection between starbursts and merging galaxies resulted in the prevailing theory that most merging galaxies go through a starburst phase (e.g. Joseph & Wright 1985; Schweizer 2005).

More recent observations have shown that merger induced starbursts are found in the minority of merging systems. These studies find that the typical increase in star-formation rate (SFR) of a merger is at most a factor of two, much lower than what would typically be considered a starburst (Ellison et al. 2013; Knapen et al. 2015; Silva et al. 2018). Work by Knapen et al. (2015) shows that the majority of galaxy mergers are found to cause a reduction in the SFR when compared to non-merging galaxies of comparable stellar masses. In total, approximately 10−20% of star-forming galaxies are found to be undergoing a merger (Luo et al. 2014; Cibinel et al. 2019) while this fraction

increases with redshift (Berrier et al. 2006; Conselice et al. 2009; López-Sanjuan et al. 2010, 2015; Lotz et al. 2011; Rodríguez-Puebla et al. 2017). Although acquiring observational evidence for the change of SFR as a function of time before and after a merger is difficult due to the long timescales involved, there is observational evidence for starbursts on the first and second close passes of two galaxies as well as coalescence. These bursts appear to last between 10^7 and 10^8 years (Cortijo-Ferrero et al. 2017). This is supported by observations that show close pairs have higher SFRs than more separated galaxies in mergers (Davies et al. 2015).

Gas rich (wet) mergers are able to support higher SFRs as there is an abundance of fuel available to create new stars (e.g. Lin et al. 2008; Perez et al. 2011; Athanassoula et al. 2016). Gas poor (dry) mergers, however, do not have gas readily available and so it is harder to form starbursts in these systems (e.g. Bell et al. 2006; Naab et al. 2006; Lin et al. 2008). As a result of dense environments containing a higher number of gas poor galaxies than gas rich galaxies, dry galaxy mergers dominate over wet mergers in dense environments (Lin et al. 2010). The fraction of dry mergers also increases with the age of the Universe (Lin et al. 2008). Due to gas poor galaxies dominating at high masses (stellar mass $\gtrsim 10^{10.7} M_\odot$), mergers of two high mass galaxies tend to be dry and, as a result, can act to suppress star formation (Robotham et al. 2014).

A study by Davies et al. (2015) find that the merger ratio of the merging galaxies also influences the SFR. In major mergers (mass ratio <3:1), the lower mass galaxy experiences a short period of enhanced star formation, while in minor mergers (mass ratio >3:1) the star formation in the lower mass galaxy is suppressed. The more massive of the two merging galaxies, however, always experiences an increase in SFR regardless of whether the merger is major or minor.

Simulations of mergers have been conducted, allowing us to study the SFRs of the merging galaxies throughout the entire merger sequence from first passage to coalescence (e.g. Springel et al. 2005; Hopkins et al. 2006; Randall et al. 2008; Rupke et al. 2010). These simulations have shown that SFR is enhanced when the merging galaxies are close to one another at first pass, second pass, and coalescence (Moreno et al. 2019). The period between first and second passes also maintains a higher SFR than in an isolated galaxy, by approximately a factor of two. This period is the majority of the merger sequence, taking approximately 2.5 Gyr of the entire 3.5 Gyr merger timescale (Moreno et al. 2019). This can explain why so few galaxies are observed in the starburst phase of a merger as the period between starbursts is much longer than the starburst period of approximately 0.5 Gyr. The starburst caused by the close passage and coalescence is also found to be stronger for head on collisions and reduces in strength as the approach of the galaxies becomes more oblique. However, the strength of a starburst is also connected to the resolution of the simulation, with lower resolution simulations finding weaker starbursts (Sparre & Springel 2016).

A major observational challenge of merger studies is the difficulty in detecting a large sample of merging galaxies. Visually identifying galaxies is time consuming and hard to reproduce; different people can classify the same galaxy differently and the same classifier may assign different labels on different days. Some of this difficulty can be reduced by employing citizen science, such as Galaxy Zoo[1] (GZ; Lintott et al. 2008), to get many members of the public to classify images of galaxies. However, such approaches are not scalable to the volume of data we expect from upcoming large surveys. Using non-parametric statistics, such as concentration, asymmetry, smoothness (CAS; e.g. Bershady et al. 2000; Conselice et al. 2000, 2003; Wu et al. 2001) or the Gini coefficient, a description of the relative distribution of flux within pixels, and the second-order moment of the brightest 20% of the light (M_{20}; Lotz et al. 2004) avoids the issues with reproducibility, especially combined with detailed galaxy merger modelling to provide a classification baseline (Lotz et al. 2010a,b). However, merger detection with these non-parametric statistics is sensitive to image quality and resolution, and suffers from a high fraction of misidentifications (Huertas-Company et al. 2015). The close pair method is also often employed, finding pairs of galaxies that are close on the sky and in redshift (e.g. Barton et al. 2000; Lambas et al. 2003; De Propris et al. 2005; Ellison et al. 2008; Rodrigues et al. 2018; Duncan et al. 2019). However, this method requires highly complete spectroscopic observations and can be contaminated with flybys (Sinha & Holley-Bockelmann 2012; Lang et al. 2014).

Deep learning has the potential to overcome some of these difficulties. Once trained, neural networks are able to perform visual like classifications of galaxies, and other astronomical objects, in a fraction of the time it takes a human, or team of humans, to classify the same objects. The classifications are also reproducible: if the same object is passed through the same neural network the result will always be the same. Deep learning techniques are becoming more commonplace in the astronomical community with uses including star-galaxy classification (e.g. Kim & Brunner 2017), galaxy morphology classification (e.g. Dieleman et al. 2015; Huertas-Company et al. 2015; Domínguez Sánchez et al. 2019), gravitational lens identification (e.g. Petrillo et al. 2017; Davies et al. 2019), and galaxy merger identification (e.g. Ackermann et al. 2018; Pearson et al. 2019).

In this work we aim to use deep learning techniques to identify merging galaxies within three data sets: the Sloan Digital Sky Survey (SDSS; York et al. 2000), the Kilo Degree Survey (KiDS; de Jong et al. 2013a,b), and the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS; Grogin et al. 2011; Koekemoer et al. 2011). These three data sets are employed so a large range of redshifts can be covered, with SDSS and KiDS at low redshift and CANDELS at high redshift. With these identifications, we compare the SFRs of the star-forming merging galaxies with the star-forming non-merging galaxies and determine if galaxy mergers have an effect on the SFR of the merging galaxies.

The paper is structured as follows. Section 2 discusses the data used and the merger selection for training our neural network. Section 3 describes the tools used in this study, including how we determined the galaxy main sequence through modelling and a brief description of the type of deep learning we employ: convolutional neural networks. This is followed by our results and discussion in Sects. 4 and 5 before we conclude in Sect. 6. Where necessary, Wilkinson Microwave Anisotropy Probe year 7 (WMAP7) cosmology (Komatsu et al. 2011; Larson et al. 2011) is adopted, with $\Omega_M = 0.272$, $\Omega_\Lambda = 0.728$, and $H_0 = 70.4 \, \mathrm{km \, s^{-1} \, Mpc^{-1}}$.

2. Data

To train the neural network, a large number of images of preclassified merging and non-merging systems are required. We also collect images of unclassified images from the same surveys to classify with our networks to increase the sample size for this study. To determine if galaxy mergers affect SFRs in the galaxies, we also require stellar masses ($M_\star$), SFR, and redshifts

---

[1] http://www.galaxyzoo.org/

for the pre-classified and unclassified galaxies. We gather these for SDSS, KiDS, and CANDELS.

These three data sets cover different redshift ranges for which merger detection is attempted: the SDSS data that we use covers the low redshift regime ($0.005 < z \leq 0.1$), along with the KiDS data ($0.00 < z \leq 0.15$), while the CANDELS data that we use goes to high redshift ($0.0 < z \leq 4.0$). The overlaps in the redshifts also allow us to examine differences due to resolution, depth, and other effects, by comparing the SDSS and KiDS results, as well as different wavelengths, by comparing the optical SDSS and KiDS with the near-infrared CANDELS. The CANDELS data also probes rest frame optical data at $z \approx 1.2$ with the three CANDELS bands used ($1.6\,\mu m$, $1.25\,\mu m$ and $814\,nm$) probing approximately the rest frame i, r, and g bands used in the SDSS data.

2.1. SDSS data release 7

For the SDSS data, we use the network trained on SDSS images from Pearson et al. (2019). The merging and non-merging galaxies used to train this network were collected following Ackermann et al. (2018). The 3003 merging galaxies are from Darg et al. (2010a,b), itself derived from classifications from the Galaxy Zoo (GZ) visual classification. These galaxies have GZ merger classification greater than 0.4 and were then visually checked again to ensure these galaxies are likely to be merging pairs. Approximately half (54%) of these merging galaxies are major mergers (Darg et al. 2010b), that is the ratio of the stellar masses of the two galaxies is less than three. For the non-merging galaxies, 3003 galaxies were randomly selected from galaxies that have their GZ merger classification less than 0.2. Cut-outs of the merging and non-merging objects were then requested from the SDSS cut-out server for data release 7[2] (DR7) to create 6006 images in the gri bands, each of 256×256 pixels and with Lupton et al. (2004) colour scaling. These images were then cropped to the central 64×64 pixels, corresponding to 25.3×25.3 arcsec or 46.5×46.5 kpc at $z = 0.1$, to reduce memory requirements while training. The merger fraction of the complete training sample, before randomly selecting the non-merging galaxies but after mass completeness cuts, is ~1.0%.

To increase the sample for analysis, all SDSS galaxies with spectroscopic redshifts between 0.005 and 0.1 were selected, to match the redshift range of the training sample, and were then classified into merging and non-merging by the Pearson et al. (2019) network, a total of 206 037 galaxies once selected for mass completeness. Again, 256×256 pixel cutouts in the gri bands were collected for these galaxies from the SDSS DR7 cutout server and the central 64×64 pixels used for classification. The $M_\star$ and SFR for these objects were then collected from the MPA-JHU catalogue[3], which uses the Kroupa (2001) initial mass function (IMF; Kauffmann et al. 2003; Salim et al. 2007; Brinchmann et al. 2004). The $M_\star$ is therefore derived from spectral energy distribution (SED) fitting while the SFR is derived from Hα observations.

For determining the galaxy main sequence (MS), the star-forming galaxies were selected by performing a cut in the $g - r$ – absolute r magnitude (M_r) plane, closely following Loveday et al. (2012), where we define star-forming galaxies as:

$$g - r < 0.08 - 0.03 M_r. \qquad (1)$$
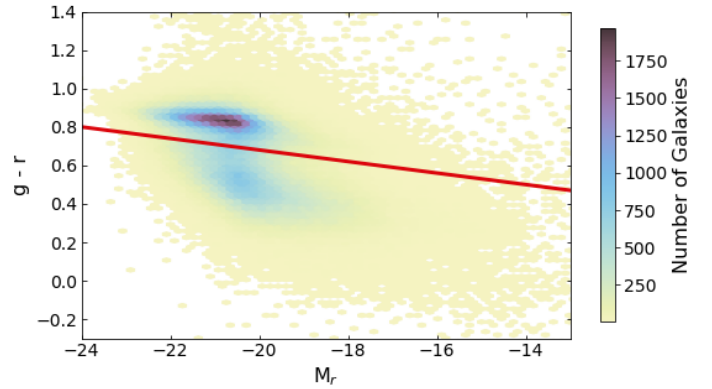
Fig. 1. Rest frame $g - r$ colour vs. absolute r magnitude (M_r) for SDSS DR7. The colour cut is shown as a red line where galaxies below the line are considered to be star-forming.

The rest frame $g - r$ colour was determined by our own fitting of the five SDSS bands with CIGALE (Noll et al. 2009; Boquien et al. 2019). A UVJ colour cut, which is used for the KiDS and CANDELS data, is not used as the wavelength coverage is not sufficient to reliably constrain the J band magnitude. A plot of this colour cut can be seen in Fig. 1. The mass limit was determined to be $\log(M_\star/M_\odot) = 10.1$, see Sect. 2.4 for details.

2.2. KiDS

For our KiDS sample, we use the latest data release 4 (DR4; Kuijken et al. 2019). We match these catalogues with the Galaxy and Mass assembly (GAMA; Driver et al. 2009) GZ catalogue (Holwerda et al. 2019; Kelvin et al., in prep.) to determine the merging and non-merging galaxies and combine this classification with non-parametric statistics (see Sect. 2.2.2). For the KiDS data, we only use r-band images to train the network, using 64×64 pixel cutouts, corresponding to 13.7×13.7 arcsec or 25.2×25.2 kpc at $z = 0.1$, and with linear colour scaling. Tests comparing multi-channel, as used with SDSS and CANDELS, and single channel images, as used with KiDS, to identify galaxy mergers have shown that using a single channel does not notably affect the results. When applying the trained CNN to unclassified objects, we use objects that lie within the GAMA09 field. This region is large enough to provide a statistically significant sample size of galaxies and has the added benefit that it has *Herschel* Spectral and Photometric Imaging Receiver (SPIRE; Griffin et al. 2010) coverage to aid with determining SFRs.

The majority of the KiDS objects in DR4 do not have estimates of physical parameters, beyond photometric redshifts (Kuijken et al. 2019). Thus, to derive $M_\star$ and SFR, we use the 9-band catalogues combined with *Herschel* ATLAS (Eales et al. 2010; Smith et al. 2017) SPIRE data de-blended with XID+ (Hurley et al. 2017; Pearson et al. 2017, see also Appendix A). From the 9-band catalogue we use the KiDS Gaussian aperture and point spread function (GAAP; Kuijken et al. 2015) flux densities for the $ugri$ optical bands and the VISTA Kilo-Degree Infrared Galaxy Survey (VIKING; Edge et al. 2013) GAAP flux densities for the $ZYJHKs$ bands, all left uncorrected for foreground extinction. SEDs are fitted to these data using CIGALE and stellar populations with a Chabrier (2003) IMF. As can be seen in Fig. 2, the $M_\star$ from CIGALE are in good agreement, within 0.2 dex on average, with those from the GAMA survey (Wright et al. 2017) estimated using the MAGPHYS
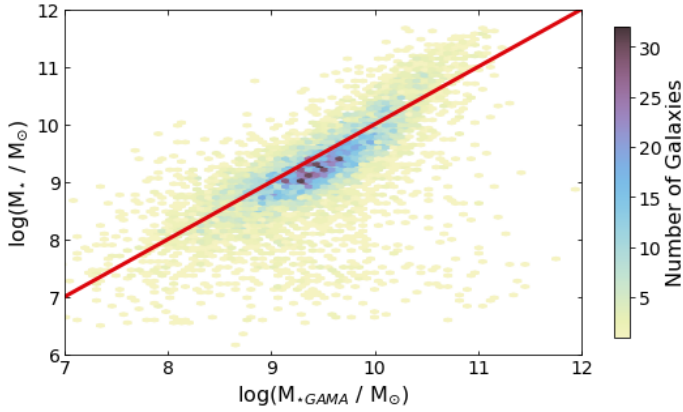
Fig. 2. Comparison of $M_\star$ from this work (y-axis) with $M_\star$ from GAMA (x-axis). The red line denotes the 1-to-1 relation. The two data sets are in reasonable agreement with the average stellar masses within 0.2 dex and remain the same with and without the inclusion of SPIRE data. The typical statistical error on $M_\star$ is 0.1 dex.
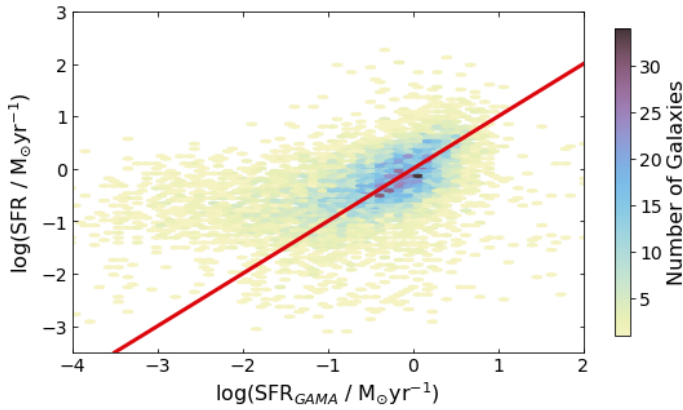

Fig. 3. Comparison of SFR from this work (y-axis) with SFR from GAMA (x-axis). The red line denotes the 1-to-1 relation. The two data sets are within 0.2 dex on average and are consistent within the typical error of 0.26 dex. Both the GAMA SFRs and the SFRs from this work are derived from SED fitting.

(da Cunha et al. 2008) SED fitting tool, which also uses the Chabrier (2003) IMF. A similar comparison is made with the SFR in Fig. 3, showing good agreement with GAMA.

To select the star-forming galaxies for determining the MS, a UVJ colour cut was employed using the rest frame $U - V$ and $V - J$ colours, determined by CIGALE during the fitting to estimate $M_\star$ and SFR, and the photometric redshifts. For this, we follow Whitaker et al. (2011):

$$
\begin{aligned}
&(U - V) > 0.88 \times (V - J) + 0.69 \quad && z < 0.5 \\
&(U - V) > 0.88 \times (V - J) + 0.59 \quad && z > 0.5 \\
&(U - V) > 1.3, (V - J) < 1.6 \quad && z < 1.5 \qquad (2) \\
&(U - V) > 1.3, (V - J) < 1.5 \quad && 1.5 < z < 2.0 \\
&(U - V) > 1.2, (V - J) < 1.4 \quad && 2.0 < z < 4.0
\end{aligned}
$$

where any galaxies that do not meet these criteria are determined to be star-forming. An example of the colour cut is shown in Fig. 4. The mass completeness limit for the KiDS galaxies was determined to be $\log(M_\star/M_\odot) = 9.6$, see Sect. 2.4 for details. This was determined using the magnitude limit from the GAMA survey of 19.8, for the r-band, as this is the limit imposed on the training sample.
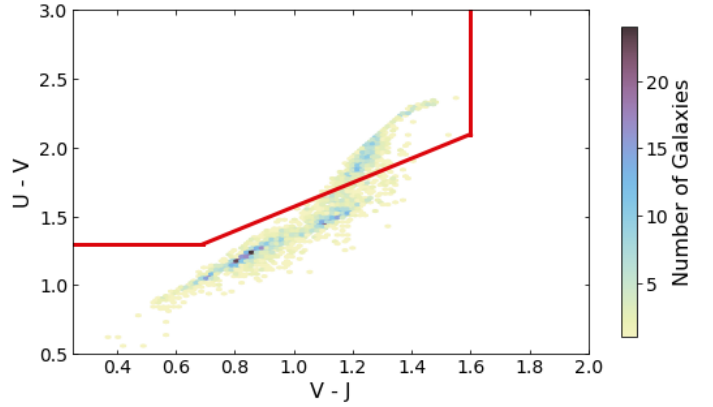
Fig. 4. Rest frame $U - V$ colour vs. rest frame $V - J$ colour for KiDS. The colour cut is shown as a red line where galaxies below and to the right of the line are considered to be star-forming.

2.2.1. KiDS-GAMA Galaxy Zoo

There are no pre-existing merger catalogues for the KiDS survey, although there are visual GZ classifications for 36 706 galaxies in the regions that overlap with the GAMA survey (KiDS-GAMA): the three GAMA equatorial fields. We can use this classification to help select a sample of merging galaxies to use with the KiDS data. As with other Galaxy Zoo (Lintott et al. 2008) projects, citizen scientists were asked to classify images of galaxies following a classification tree, as described in Holwerda et al. (2019), through the GZ web interface[4] and we use the vote fractions that are weighted for user performance. These weighted vote fractions have votes from users that frequently disagree with the majority of other users weighted lower, reducing their influence on the overall vote fraction. These galaxies were selected to have redshifts between 0.002 and 0.15 and GAMA data quality flags are used to ensure only science targets are shown. Of interest here is the question concerning galaxy interactions. This question asks the classifier to identify merging galaxies, galaxies with tidal tails, galaxies that are both merging and have tidal tails or galaxies that show neither of these features. The latter of these classifications, galaxies that have neither tidal tails nor show evidence of a merger, is what is used here to help identify galaxy mergers and will hence forth be referred to as merger_neither_frac. Galaxies that have merger_neither_frac less than 0.5, that is less than half the people who classified the galaxy thought it showed no tidal features or merger indications, is used here to for the basis of the merging galaxy sample with further refinements added.

2.2.2. KiDS merger selection

The visual GZ merger classifications require validation with other methods, as chance projections or star-galaxy overlaps can be misidentified as merging galaxies (Darg et al. 2010a,b). To do this, we use the Gini, the second-order moment of the brightest 20% of the light (M_{20}), concentration (C), asymmetry (A) and smoothness (S) non-parametric parameters (Lotz et al. 2004; Bershady et al. 2000; Conselice et al. 2000, 2003; Wu et al. 2001). For each of the galaxies in the KiDS-GAMA sample, we derive these five non-parametric statistics using the python code statmorph (Rodriguez-Gomez et al. 2019).

---

[4] http://www.galaxyzoo.org/
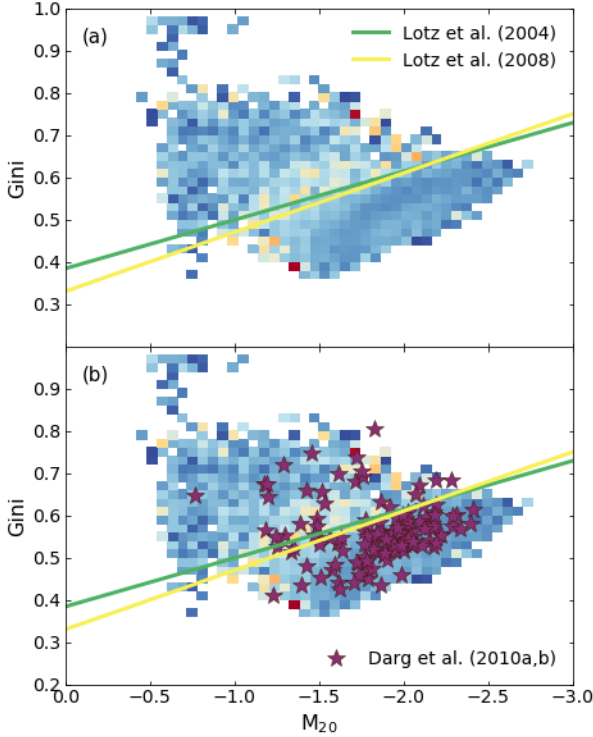
Fig. 5. Gini vs. M_{20} for KiDS-GAMA GZ galaxies binned by Gini and M_{20}. The average `merger_neither_frac` from GZ within each bin is shown from low (red) to high (blue). The green line is the Lotz et al. (2004) split between merging and non-merging galaxies while the yellow line is the Lotz et al. (2008) split. Regions with low `merger_neither_frac` are visually identified as merging galaxies. *Panel b* includes the visually confirmed mergers from Darg et al. (2010a,b) as purple stars.

Fig. 6. Asymmetry (A) vs. Smoothness (S) for KiDS-GAMA GZ galaxies binned by A and S. The average `merger_neither_frac` from GZ within each bin is shown from low (red) to high (blue). Regions with low `merger_neither_frac` are visually identified as merging galaxies. The orange line denotes the Conselice (2003) split between merging and non-merging galaxies. *Panel b* includes the visually confirmed mergers from Darg et al. (2010a,b) as purple stars.

There has been found to be a division between merging and non-merging galaxies using the Gini and M_{20} statistics: Lotz et al. (2004) found that galaxies can be considered to be non-mergers if
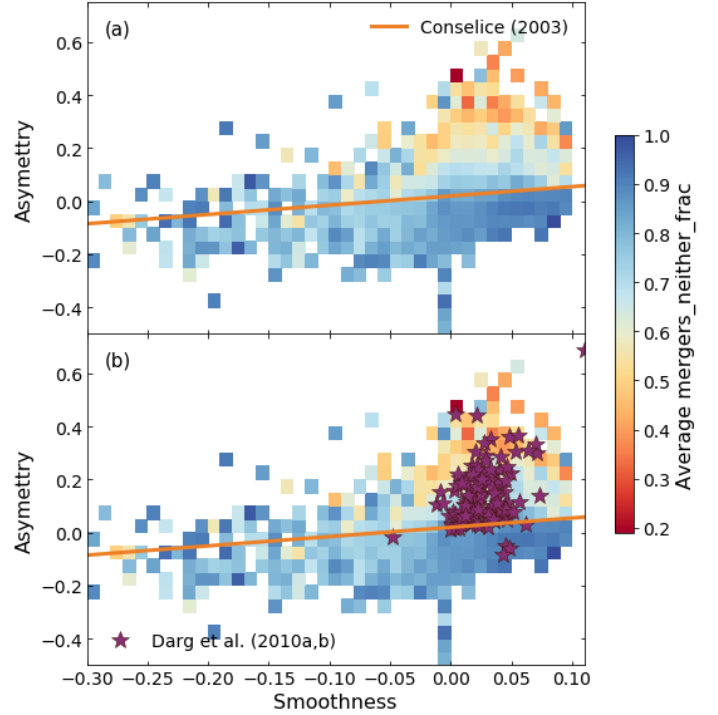
$$\text{Gini} < -0.115 M_{20} + 0.384 \qquad (3)$$

while Lotz et al. (2008) found a similar result with non-mergers defined as

$$\text{Gini} < -0.15 M_{20} + 0.33. \qquad (4)$$

We also populate the Gini-M_{20} parameter space, bin by Gini and M_{20}, and show the average `merger_neither_frac` of the galaxies inside each bin, as seen in Fig. 5. The `merger_neither_frac` is the fraction of GZ votes that say the galaxy has no indication of a galaxy merger or tidal tails. In doing this, we find that galaxies found to be mergers in the KiDS-GAMA GZ typically lie on or above these two lines. However, as can be seen in Fig. 5, there are also a large number of galaxies that lie above these lines that are classified by GZ as non-mergers: the merging galaxies appear to form a valley. Overlaying the visually confirmed merging galaxies from Darg et al. (2010a,b) that fall within the KiDS coverage, we also find that the majority of these galaxies lie below the Lotz et al. (2004, 2008) lines, as can be seen in Fig. 5b, suggesting that this is a poor choice to determine merger status for this KiDS data set.

This disparity may be a result of the different data used. The Gini and M_{20} statistics are calculated from the images and so depend on the resolution and signal-to-noise of the images (Lotz

et al. 2004). The flux distribution of a lower resolution image will be different, the same flux will be spread across fewer pixels in a lower resolution images as well as removing smaller scale structures, which will increase the uncertainties in these statistics. Similarly, higher signal-to-noise images will reveal fainter features of a galaxy that will also affect the Gini and M_{20}. The Gini and M_{20} have been found to be reasonably consistent when the signal-to-noise is above two but M_{20} is particularly sensitive to resolution (Lotz et al. 2004). The data used in Lotz et al. (2004) is lower resolution than KiDS while Lotz et al. (2008) uses *Hubble* Space Telescope data with higher resolution.

If instead we use the asymmetry (A) and smoothness (S) statistics, which have been found to be not overly sensitive to resolution as M_{20} (Lotz et al. 2004), we find a merging sample that agrees much better with the visual classification. It has been found, by Conselice (2003), that the merging galaxies lie above

$$A = 0.35 S + 0.02. \qquad (5)$$

As can be seen in Fig. 6, this classification is in good agreement with the visual classifications from GZ. Overlaying the Darg et al. (2010a,b) mergers, we find that the majority lie above Eq. (5). Based on this agreement, we select our merging sample to be those galaxies that have `merger_neither_frac` from GZ less than 0.5 and lie above Eq. (5), with non-merging galaxies defined as those with `merger_neither_frac` greater than 0.5 and lie below Eq. (5). This results in 1917 merging galaxies that we use to train the KiDS network. By matching these galaxies to the nearest galaxy within 3 arcsec in the full GAMA catalogue (Wright et al. 2017) and selecting the pairs that have redshifts within 0.05, we find that approximately half of these

galaxies (6 of 14) are major mergers. The total number of matched pairs is very low, and misses pairs where the secondary galaxy is below the magnitude limit of the survey, but this fraction is in line with that seen by Darg et al. (2010b) in the SDSS data. We randomly select a further 1917 galaxies from the 20 842 that lie below Eq. (5) and have `merger_neither_frac` greater than 0.5 to form the non-merging sample. With these classifications for merging and non-merging galaxies, and after mass completeness cuts, the merger fraction of the GZ galaxies is 8.4%.

2.3. CANDELS

To train the CANDELS network, we use the visual classifications for the Great Observatories Origins Deep Survey – South (GOODS-S; Giavalisco et al. 2004) from Kartaltepe et al. (2015). This catalogue contains galaxies with H magnitude less than 24.5 that have been classified by a small number of professional astronomers and we select objects with photometric redshift below 4.0. Of interest to this work are the classifications that identify mergers (`merger`), interaction within a segmentation map (`Int1`), interaction with a galaxy outside of the segmentation map (`Int2`), a non interacting companion (`Comp`) or no interaction (`NoInt`). During the classification, only one of these identifications may be chosen. The catalogue also contains an `Any_Int` category, which combines the `merger`, `Int1`, and `Int2` identifications.

We define galaxies as merging if the `Any_Int` classification is greater than 0.6 (that is more that 60% of people believe that the galaxy is interacting) and we define galaxies as non-merging if the `Any_Int` classification is less than 0.5. As with the KiDS galaxies, we match the merging galaxies to the rest of the CANDELS catalogue within 3 arcsec and selecting the pairs that have redshifts within 0.05, we find that approximately half of these galaxies (4 of 9) are major mergers. Again, the total number of matched pairs is very low, and this method misses pairs where the secondary galaxy is below the magnitude limit of the survey, but this fraction is in line with that seen in the SDSS data. Cutouts for these objects were created from the 1.6 μm, 1.25 μm, and 814 nm images. As the 814 nm images are twice the angular resolution of the other two bands, these images are reduced in size by averaging the flux density in 2×2 pixel groups. The 1.6 μm, 1.25 μm, and 814 nm bands are then used as the red, green, and blue channels in the images, with simple linear colour scaling. As with the SDSS and KiDS images, the CANDELS images are 64×64 pixels, corresponding to 3.8×3.8 arcsec or 32.7×32.7 kpc at $z = 1.5$. Objects with clear artefacts within the image were removed. This resulted in 694 merging galaxies and we randomly select a further 694 from the 4428 non-merging galaxies that meet our criteria. The merger fraction for the training sample using these criteria, and after mass completeness cuts, is 15.5%.

To increase the CANDELS sample for our analysis, we classified all CANDELS galaxies with H-magnitude <24.5 and redshift between 0.0 and 4.0, to match the training sample, from the Cosmic Evolution Survey (COSMOS; Scoville et al. 2007), Extended Groth Strip (EGS; Davis et al. 2007), UKIRT Infrared Deep Sky Survey (UKIDSS) Ultra-Deep Survey (UDS; Lawrence et al. 2007; Cirasuolo et al. 2007) fields with the CANDELS network (once trained). Images for these galaxies were created as above. The H-magnitude and SED derived SFR and $M_\star$, assuming a Chabrier (2003) IMF, come from Guo et al. (2013), Santini et al. (2015) for GOODS-S, Nayyeri et al. (2017) for COSMOS, Stefanon et al. (2017) for EGS, and Santini et al. (2015) for UDS. As these catalogues contain a number of
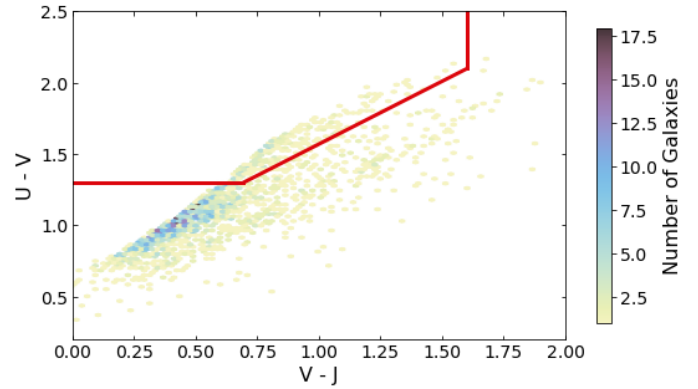
Fig. 7. Rest frame $U - V$ colour vs. rest frame $V - J$ colour for CANDELS-z000. The colour cut is shown as a red line where galaxies below and to the right of the line are considered to be star-forming.

different $M_\star$ and SFR values, the "M_med" is used for $M_\star$ and we average "SFR_11a_tau", "SFR_13a_tau", "SFR_2a_tau", "SFR_14a", "SFR_14a_const", "SFR_14a_deltau", "SFR_14a_lin", "SFR_14a_tau", "SFR_6a_deltau", "SFR_6a_invtau", and "SFR_6a_tau" for SFR, as these columns are common across all catalogues. These different SFR values assume different star-formation histories (SFH) where "cons" is a constant SFH, "tau" is an exponentially declining SFH, "deltau" is a delayed-exponential, "lin" is linearly increasing, and "inctau" is exponentially increasing. The numbers refer to the investigator within the CANDELS team who lead the determination of that SFR (Stefanon et al. 2017). For the redshift, the "z_best" value in the catalogues were used. This value is the spectroscopic redshift, if available, or the best photometric redshift from six members within the CANDELS team (Guo et al. 2013; Santini et al. 2015; Nayyeri et al. 2017; Stefanon et al. 2017).

To determine which CANDELS galaxies are star-forming, we again apply the UVJ colour cuts defined in Eq. (2) using the rest frame $U - V$ and $V - J$ colours in the CANDELS catalogues, as shown in Fig. 7. Mass completeness limits were calculated to be $\log(M_\star/M_\odot) = 8.3$, 8.7, 9.1, 9.4, and 9.9 within redshift bins with edges at $z = 0.0$, 0.6, 0.85, 1.21, 1.66, and 4.0, see also Sect. 2.4 below. These redshift bins were selected so there are approximately 2000 galaxies within each bin after cutting for mass completeness. For ease of reference, these redshift bins shall be referred to as CANDELS-z000, CANDELS-z060, CANDELS-z085, CANDELS-z121, and CANDELS-z166. A summary of all data sets is presented in Table 1.

2.4. Mass completeness

Mass completeness limits were determined empirically by following Pozzetti et al. (2010) and using the galaxies identified as star-forming. For each galaxy, the mass the galaxy would need to have to be detected at the magnitude limit (M_{lim}) was calculated with

$$\log(M_{lim}) = \log(M) - 0.4(x_{lim} - x), \qquad (6)$$

where x is the observed magnitude in the r-band (for SDSS and KiDS) or H-band (for CANDELS) and x_{lim} is the limiting magnitude of the observation. The limiting magnitudes for SDSS and CANDELS are 17.77 and 24.50 respectively. The KiDS limiting magnitude is 19.8, the limit of the GAMA survey. The faintest 20% of objects were selected and the limiting mass was the M_{lim} value that 90% of these faintest objects lie below. This was done as a function of redshift by binning the galaxies into redshift bins

Table 1. Summary of data used.

Data	Resolution (arcsec)	Magnitude limit	Redshift range	Mass limit $\log(M_\star/M_\odot)$	Training sample per class	Complete sample (Mass limited)
SDSS	1.4	17.77	$0.005 < z \leq 0.1$	10.1	3003	206 037
KiDS	0.77	19.8 [a]	$0.00 < z \leq 0.15$	9.6	1917	1270
CANDELS-z000	0.15	24.5	$0.00 < z \leq 0.60$	8.3	694 [b]	2072
CANDELS-z060	0.15	24.5	$0.60 < z \leq 0.85$	8.7	694 [b]	2004
CANDELS-z085	0.15	24.5	$0.85 < z \leq 1.21$	9.1	694 [b]	2031
CANDELS-z121	0.15	24.5	$1.21 < z \leq 1.66$	9.4	694 [b]	2010
CANDELS-z166	0.15	24.5	$1.66 < z \leq 4.00$	9.9	694 [b]	1910

Notes. The SDSS and KiDS limiting magnitudes are in *r*-band while the CANDELS limiting magnitude in *H*-band. [a] As the training set is derived from GAMA classifications, the limiting magnitude is that of the GAMA survey not that of the KiDS survey. [b] The CANDELS network was trained with 694 galaxies per class for galaxies with $0.00 < z \leq 4.00$. The galaxies were split into the redshift bins shown after classification.

as described in Sect. 4 below. These completeness limits were then applied to the entire galaxy population.

3. Tools

3.1. Convolutional neural networks

Convolutional neural networks (CNNs) are a subset of deep learning (e.g. Lecun et al. 2015, and references therein). CNNs are used for image classifications and employ a series of non-linear mathematical functions, known as neurons, each with a weight and bias value. The structure of a CNN is built from a number of layers of these neurons. The lower layers are created from two-dimensional kernels that are convolved with the output of the layer below, giving CNN its name. Upper layers are one-dimensional and each neuron in these layers is connected to every neuron in the layer below. Forming a network in such a way can rapidly create a large number of neurons that require training resulting in many more free parameters within the network than there are data to train them. To reduce this dimensionality, pooling layers are employed between the lower convolutional layers. These pooling layers group the inputs into it and pass on the maximum or average value of the group, depending on the type of pooling used, with the grouping done in two-dimensions. The result is an output that is smaller in the width-height plane but has the same depth as the input. The weights and biases of the neurons within a network are trained, in the case of supervised learning used here, by passing labelled data through the network and requiring the output classification to converge on these labels. A complete and thorough description of CNNs is beyond the scope of this paper but further details are explained in Lecun et al. (1998).

This paper uses the definitions of Pearson et al. (2019) for the terms to describe the properties of CNNs. These terms may be an alternate nomenclature to other works or may be unfamiliar. To avoid confusion we reproduce these definitions in Appendix B.

3.1.1. Architecture of the CNN

For this work, we use the architecture developed in Pearson et al. (2019) and use this to train on data from CANDELS and KiDS (the network trained in Pearson et al. 2019 is used on the SDSS images). This network is built with `Tensorflow` (Abadi et al. 2015) and comprises of a series of four, two-dimensional convolutional layers followed by two one-dimensional, fully connected layers of 2048 neurons. The convolutional layers have 32, 64, 128, and 128 kernels of 6×6, 5×5, 3×3, and 3×3 pixels for the first, second, third, and fourth layers respectively with

the stride, how far the kernel is moved as it scans the input, set at 1 pixel for all layers and the zero padding is set to "same" to pad each edge of the image with zeros evenly (if required). 2×2 pixel max-pooling is applied after the first, second, and fourth convolutional layers to reduce the dimensionality of the network. Batch normalisation (Ioffe & Szegedy 2015) is applied after each layer, scaling the output between zero and one, and we use Rectified Linear Units (ReLU; Nair & Hinton 2010) for activation, which returns $\max(x, 0)$ when passed x. We also apply dropout (Srivastava et al. 2014) after each activation, to help reduce over-fitting, with a dropout rate of 0.2 to randomly set the output of neurons to zero 20% of the time during training. The loss of this network is determined using softmax cross entropy and optimised using the Adam algorithm (Kingma & Ba 2015) with a learning rate of 5×10^{-5}. The architecture can be seen schematically in Fig. 8.

The inputs to the network are 64×64 pixel images with either three (CANDELS) or one (KiDS) colour channels, that are globally scaled between 0 and 1, preserving the relative flux densities for multi colour images. The output layer has two neurons, one for each of the merging and non-merging classes, and uses a softmax output, providing the probability for each class in the range [0, 1] that sum to unity, that is softmax maps the un-normalised input into it to a probability distribution over the output classes if the training, test, and validation data sets have an equal number of each class. In the following, we use the output for the merger class (`frac_merger`) although this can be considered to be equivalent to using the output for the non-merger class as it is 1-(`frac_merger`) in our binary classification.

Due to the limitations of this architecture, specifically the use of fully connected layers, it is not possible to have input images of different pixel sizes. As a result, it is not possible to use cutouts of different sizes that maintain the relative size of the galaxy within the image. It is possible to resize the images but this risks losing small scale structure when downscaling galaxies or creating artefacts when upscaling. This may cause issues for large galaxies at low redshifts and low resolution as the primary galaxy may fill the image. However, it is possible to correctly identify the galaxies that fill, or are larger than, the image if the training set contains these types of galaxies, as has been shown in Pearson et al. (2019).

3.1.2. Training, validation, and testing

If there are an unequal number of images in the two classes, the larger class size is reduced by randomly removing images until the classes are the same size. The images were then subdivided
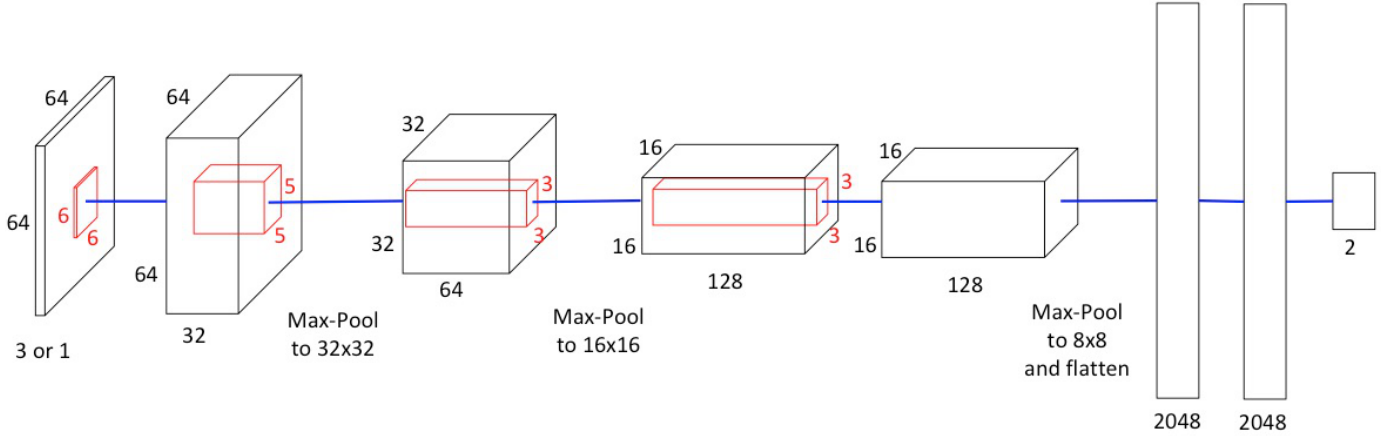
Fig. 8. Schematic representation of the architecture of the CNN used with input (three or one colour, 64×64 pixel image) on the left and output (binary classification of merger or non-merger) on the right. The sizes of the kernels (red) and layers are shown with the input layer having a depth of three for the SDSS (*gri*) and CANDELS ($1.6\,\mu$m, $1.25\,\mu$m, and 814 nm) data and one for the KiDS (*r*-band) data.

into three groups: 80% were used for training, 10% for validation, and 10% for testing. The training set was the set used to train the network while the validation was used to see how well the network was performing as training progressed. The testing set was used once, and once only, to test the performance of the network deemed to be the best from the validation. Testing images are not used for validation to prevent accidental training on the test data set.

During training, the input images were augmented. To increase the number of galaxies to train with for CANDELS and KiDS, the images were randomly augmented in one of the following ways: the image was translated by up to 4 pixels in each direction, zoomed by up to 2×, rotated by a random angle about the galaxy, skewed by a random angle between 10° and 30° or passed without change. For the first four of these options, random Gaussian noise was added so the galaxy appeared "new" to the network. These augmented images were then randomly rotated by 0°, 90°, 180° or 270° to reduce sensitivity to galaxy orientation. Once complete, we have one network for the CANDELS data and one for the KiDS data.

3.2. Forward modelling of the galaxy main sequence

In determining if galaxy mergers have an effect on the SFRs of galaxies, it is necessary to remove the mass dependence from the SFR. To do this we correct for the slope of the main sequence of star-forming galaxies (MS, e.g. Brinchmann et al. 2004; Noeske et al. 2007; Elbaz et al. 2007; Speagle et al. 2014; Pearson et al. 2018). The MS is an observed tight correlation between the $M_\star$ and SFR of star-forming galaxies that exists out to at least $z = 6$ with a scatter of ~0.3 dex that is mass and redshift independent. The slope of the MS is found to be less than one and to depend on the redshift (Pearson et al. 2018). To correct for this slope, it is necessary to first determine the MS of the galaxies.

To determine the MS, we follow Pearson et al. (2018) and use forward modelling. Assuming a linear MS, we use the Markov Chain Monte-Carlo (MCMC) sampler `emcee` (Foreman-Mackey et al. 2013) to simultaneously fit the slope, normalisation, and scatter of all the galaxies that are star-forming. At each sampled point in the parameter space, model SFRs are created using the observed stellar masses, observed redshifts, and the corresponding positions along the sampled MS. These SFRs are then perturbed by selecting a random number from a Gaussian distribution, with the standard deviation equal to the sampled

scatter, truncated to match the observed upper and lower SFR limits. To include observational uncertainties on SFR and $M_\star$, both the simulated SFR and the observed $M_\star$ of each galaxy are perturbed again by a random number sampled from a Gaussian distribution with the standard deviation equal to the error in the observed SFR or $M_\star$.

At each step, the mock SFR-$M_\star$ plane is compared to the observed SFR-$M_\star$ plane by binning the data into identical bins in $M_\star$. The means and standard deviations of the SFRs inside each bin are calculated and the results from the mock data are compared to the results from the observed data. The closer the means and standard deviations in the mock bins are to the observed bins, the more likely the model is a correct representation of the observed data. Due to the large number of objects, for the SDSS data we randomly select 50 000 star-forming objects within each redshift bin to determine the MS.

4. Results

4.1. Performance of the CNN

The network we use for the SDSS objects is that of Pearson et al. (2019). For completeness, we repeat these results here. The SDSS network achieves an accuracy of 0.932 at validation with a cut threshold of 0.5, that is any object with `frac_merger` > 0.5 is classified as a merging galaxy. We can alter the threshold value to find the threshold value that simultaneously minimises the fall out and maximises the recall. By doing this to set the threshold to 0.57, the accuracy increases to 0.935. Using the same threshold of 0.57 at testing, the final accuracy of the SDSS network is 0.915, with recall, precision, specificity, and negative predictive value (NPV) of 0.920, 0.911, 0.910, and 0.919 respectively. We apply the network to 256 497 SDSS galaxies with spectroscopic redshifts between 0.005 and 0.1, to match the training set, resulting in 28 971 ± 2578 (14.1 ± 1.3%) galaxies being identified as mergers. The errors in these merger counts are derived from the precision of the network, see Appendix B for the definition. The number of merging galaxies is multiplied by the precision and the difference between this value and the original count is taken as the error for the number of mergers and number of non-mergers. This is likely an underestimate as the precision assumes equal population sizes of mergers and non-mergers, which is evidently not the case.

Table 2. Statistics for trained CNNs.

| | SDSS [a] | KiDS | CANDELS |
	$0.005 < z < 0.1$	$0.00 < z \leq 0.15$	$0.00 < z \leq 4.0$
Cut threshold	0.57	0.52	0.47
ROC area	0.966	0.957	0.861
Recall	0.920	0.942	0.870
Precision	0.911	0.874	0.789
Specificity	0.910	0.864	0.768
NPV	0.919	0.938	0.855
Accuracy	0.915	0.903	0.818

Notes. Definitions of terms can be found in Appendix B. [a]The SDSS network is that of Pearson et al. (2019).

For the KiDS network, we use the CNN to identify galaxies that fall within the GAMA09 field. This network achieves an accuracy of 0.942 at validation with a cut threshold of 0.5. If we alter the threshold to 0.52, to simultaneously minimise the fall out and maximise the recall, the accuracy increases to 0.948. Using the same threshold of 0.52 at testing, the final accuracy of the KiDS network is 0.903, with recall, precision, specificity, and NPV of 0.942, 0.874, 0.864, and 0.938 respectively. Despite the resolution of the KiDS images being higher than that of the SDSS images, the same image size resulted in the best performance for the KiDS network: 64×64 pixels. Larger images were tried but these networks did not perform as well. Applying the KiDS network to all galaxies in the GAMA09 field with photometric redshifts below 0.15, a total of 1270 galaxies, we identify 436 ± 55 ($30.0 \pm 4.3\%$) merging galaxies.

The CANDELS network achieves an accuracy of 0.826 at validation with a cut threshold of 0.5. If we decrease the threshold to 0.47, the accuracy increases to 0.840. Using the same threshold of 0.47 at testing, the final accuracy of the CANDELS network is 0.818, with recall, precision, specificity, and NPV of 0.870, 0.789, 0.768, and 0.855 respectively. The poorer results for the CANDELS network is likely due to fewer preclassified objects to train the network with, 694 per class for CANDELS compared to 3003 for SDSS, as well as the higher redshifts of the training objects. The CANDELS images also cover a much larger redshift range, resulting in a greater distribution of sizes in the image for galaxies at the same mass than the SDSS images. Ideally, it would be preferable to split the galaxies into redshift bins and train a network per redshift to minimise this effect, however with so few objects it is not feasible. We apply the CANDELS network to the objects with H-magnitude <24.5 and $0.0 < z < 4.0$ in the CANDELS COSMOS, EGS, and UDS fields and identify 3535 ± 746 merger candidates out of the 10 027 galaxies in these three fields. This is a merger fraction of $35.3 \pm 7.4\%$, which is high. The statistics for all the networks are presented in Table 2 and examples of non-merger and mergers selected by the CNNs can be found in Appendix C.

4.2. SDSS

To determine the effect of galaxy mergers on SFR, we determine the effect of mergers on the MS subtracted SFR. We fit the MS to all the star-forming galaxies, both mergers and non-mergers together, and the MS we have fitted to the SDSS data is shown overlaid onto all the non-merging and merging galaxies in Fig. 9. The MS subtracted SFR of the merging and non-merging galaxies are then compared by fitting a skewed Gaussian distribution,
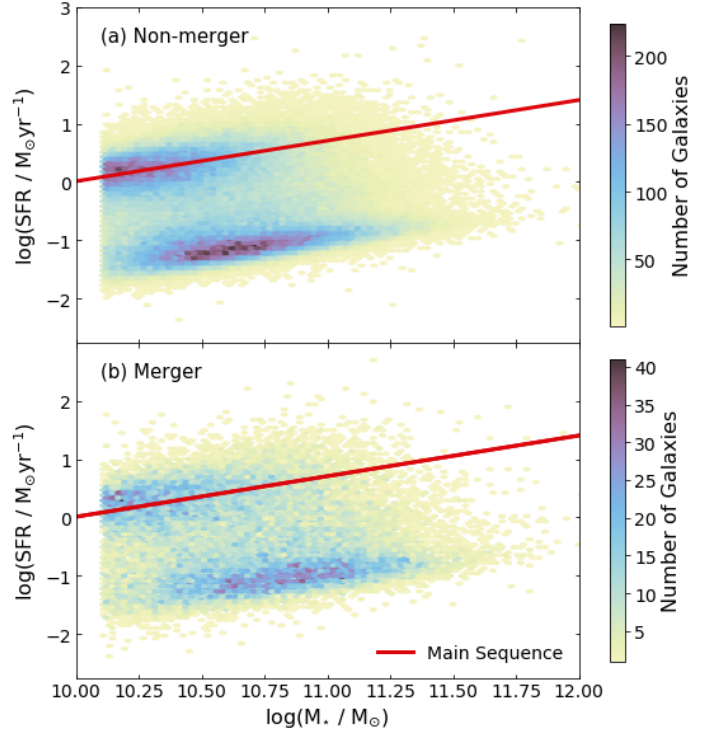
Fig. 9. SFR-$M_\star$ plane populated with (a) non-merging galaxies and (b) merging SDSS galaxies. The colour indicates the number density from low (light yellow) to high (dark purple). Overlaid in red is the MS that has been fitted to all star-forming galaxies. As can be seen, the distributions of the merging and non-merging galaxies are similar with respect to the plotted MS.

of the form

$$y = \frac{A}{\sigma} \exp\left(\frac{(x - \mu)}{2\sigma^2}\right)\left(1 + \text{erf}\left[\frac{\alpha(x - \mu)}{\sqrt{2}\sigma}\right]\right), \quad (7)$$

to the distributions of the merging and non-merging galaxies, where A is the amplitude, μ, and σ are the mean and standard deviation of the Gaussian, α is the description of skewness, and erf is the error function.
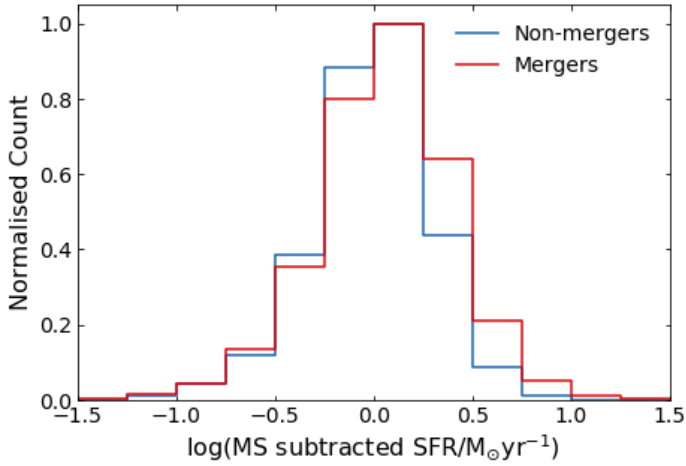
To fit the skewed Gaussian we bin the MS subtracted SFR with bin sizes of 0.25 dex, between -3 $\log(M_\odot \text{ yr}^{-1})$ and 2 $\log(M_\odot \text{ yr}^{-1})$ and fit the skewed Gaussian to the number of galaxies in each bin. The errors on these counts were determined by generating 100 realisations of the MS subtracted SFR by perturbing the SFR and $M_\star$ of each galaxy by a random number drawn from a Gaussian distribution centred on the observed SFR or $M_\star$ and with the error on the value as the standard deviation. Each realisation was then binned in the same way as the observations and the standard deviation of the counts in the bins of the 100 realisations were taken as the errors on the counts of the observations. The `scipy.optimize` package `curve_fit` was then used to fit the skewed Gaussian distribution to the counts in the bins and account for their errors. The distributions are presented in Fig. 10 with the parameters for the skewed Gaussian fits in Table 3.

Comparing the skewed Gaussian fits to the distributions, we find that the mean for the star-forming mergers and non-mergers are consistent within three times the error of the mean (σ_μ) and the merging galaxies have higher mean MS subtracted SFR. This suggests that the star-forming population has a slightly, but not significantly, increased SFR when undergoing a merger.

Table 3. Best fit parameters for skewed Gaussian distributions fitted to star-forming SDSS data.

Parameter	Merger	Non-merger
μ	0.33 ± 0.02	0.25 ± 0.01
σ	0.43 ± 0.02	0.39 ± 0.01
α	-1.29 ± 0.18	-1.52 ± 0.16

Notes. μ and σ are in units of $\log(M_\odot \, \mathrm{yr}^{-1})$.

Table 4. Best fit parameters for skewed Gaussian distributions fitted to star-forming KiDS data.

Parameter	Merger	Non-merger
μ	0.44 ± 0.1	0.13 ± 0.18
σ	0.47 ± 0.09	0.41 ± 0.08
α	-1.53 ± 0.91	-0.63 ± 0.78

Notes. μ and σ are in units of $\log(M_\odot \, \mathrm{yr}^{-1})$.

Fig. 10. Distribution of MS subtracted SFR for star-forming SDSS non-merging galaxies (blue) and merging galaxies (red). As can be seen, the merging star-forming population has a slightly higher mean MS subtracted SFR.
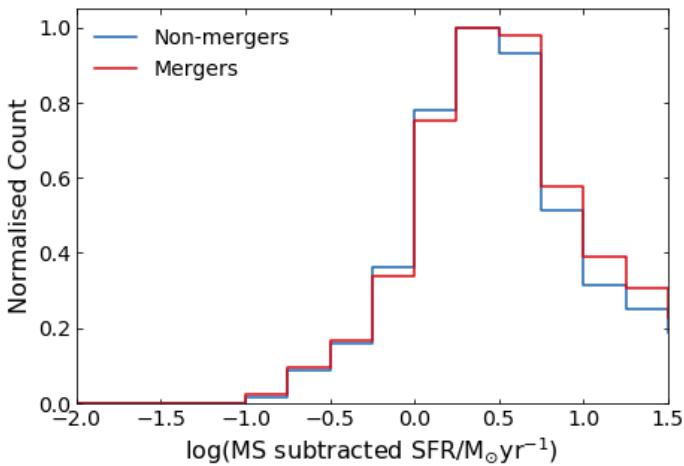
Fig. 11. Distribution of MS subtracted SFR for star-forming KiDS non-merging galaxies (blue) and merging galaxies (red). As can be seen, the merging star-forming galaxies have a similar mean MS subtracted SFR to the non-merging galaxies.

4.3. KiDS

As with the SDSS data, we fit a skewed Gaussian distribution to the MS subtracted SFR of the star-forming galaxies. An example of the resulting MS subtracted SFR distributions for the KiDS is shown in Fig. 11. Table 4 shows that the merging star-forming galaxies have higher average SFRs. The differences not large, with the mean MS subtracted SFR being within $3\sigma_\mu$ of each other.

Fig. 12. Distribution of MS subtracted SFR for the $0.85 < z \leq 1.21$ redshift bin for CANDELS non-merging galaxies (blue) and merging galaxies (red). This is the only data that is fitted with a double Gaussian distribution due to the clear multi-modal population. As can be seen, the main and secondary populations have a slightly higher mean MS subtracted SFR than the non-merging galaxies.

4.4. CANDELS

Due to the larger redshift coverage of the CANDELS data, we can examine if the impact of galaxy mergers on SFR changes as a function of redshift. To do this, we divided the data into redshift bins with edges at $z = 0.0$, 0.6, 0.85, 1.21, 1.66, and 4.0, each with its own mass completeness limit and containing approximately 2000 galaxies after mass completeness cuts have been applied. Each redshift bin also had its own main sequence fitted as outlined in Sect. 3.2. For ease of reference, these redshift bins shall be referred to as CANDELS-z000, CANDELS-z060, CANDELS-z085, CANDELS-z121, and CANDELS-z166.

As before, we fit the distributions of the star-forming CANDELS galaxies with a skewed Gaussian function. However, there is an indication of a second, high SFR population in CANDELS-z085 ($0.85 < z \leq 1.21$), identifiable when the error on the skew of the both the merging and non-merging distributions are greater than 10^4 and so in that bin only, we fit a double Gaussian distribution and consider the lower mean to be the mean of the star-forming population. The distributions for the MS subtracted SFR for this redshift bin is shown in Fig. 12.

Using the best fitting values for the skewed and double Gaussian functions presented in Table 5, in the lowest redshift bin the merging galaxies act to suppress the SFR of galaxies, with the mean MS subtracted SFR for merging galaxies lower than non-merging galaxies by more than $3\sigma_\mu$. At redshifts above $z = 0.60$, the star-forming mergers have a higher mean than the star-forming non-mergers or are consistent within $3\sigma_\mu$.

Table 5. Best fit parameters for skewed or double Gaussian distributions fitted to star-forming CANDELS data.

Redshift	Parameter	Merger	Non-merger
	μ	-0.37 ± 0.17	-0.27 ± 0.02
$0.0 < z \le 0.6$	σ	0.47 ± 0.08	0.29 ± 0.02
	α	9999.489 ± 87523220.85	1.21 ± 0.26
	μ	0.10 ± 0.12	0.02 ± 0.11
$0.6 < z \le 0.85$	σ	0.24 ± 0.07	0.24 ± 0.06
	α	-0.878 ± 1.17	-0.79 ± 0.99
$0.85 < z \le 1.21$	μ	-0.12 ± 0.01	-0.16 ± 0.01
	σ	0.20 ± 0.01	0.20 ± 0.01
	μ	-0.43 ± 0.02	-0.42 ± 0.02
$1.21 < z \le 1.66$	σ	0.62 ± 0.03	0.52 ± 0.03
	α	5.058 ± 1.24	3.13 ± 0.69
	μ	-0.47 ± 0.02	-0.5 ± 0.02
$1.66 < z \le 4.0$	σ	0.65 ± 0.03	0.61 ± 0.03
	α	3.241 ± 0.51	3.7 ± 0.61

Notes. For the $0.85 < z \le 1.21$ bin, where a double Gaussian is used, the star-forming component is the component with the lowest μ. μ, and σ are in units of $\log(M_\odot \, \text{yr}^{-1})$.

Fig. 13. Average MS subtracted SFR of star-forming galaxies for SDSS merging (purple circle) and non-merging (dark blue diamond); KiDS merging (light blue circle) and non-merging (green diamond); and CANDELS merging (orange circles) and non-merging (red diamonds) galaxies. As can be seen, the change in SFR between the merging and non-merging galaxies is typically small.

5. Discussion

Here we present discussions of our results. We note that direct comparisons between the results of the three data sets is difficult due to the different definitions of mergers employed for the training data sets as well as difference in data quality, such as depth and resolution, which can also influence merger identification. While the merger definitions are similar, as they are all based on visual classification, the specifics of the definitions differ. The classifications also cover both major and minor mergers, with approximately half of each training set comprising of major mergers. This likely results in a similar split for the mergers classified by our networks.

5.1. Merger influence on SFR

Across the SDSS, CANDELS, and KiDS data sets there is a difference between the SFRs of the merging and non-merging galaxies. However, the difference between the two is small and varies between the data sets as well as within the data sets. What is evident is that the merging systems are not only found as starburst galaxies but also as star-forming and quiescent systems.

Comparing the SDSS data with the KiDS data, we find little difference in how mergers are affecting the SFR. Both data sets show that star-forming merging galaxies have a slight increase in SFR. Within the CANDELS-z000 data the opposite is found: we find that there is a decrease in the MS subtracted SFR, suggesting that galaxy mergers are acting to reduce the SFR of the star-forming galaxies. The full comparison between the average SFRs for all data sets at all redshifts studied can be seen in Fig. 13.

The slight difference between the merging and non-merging SFRs is also not a result of the observation bands or methods used to derive the SFR. The mergers in all three surveys are detected using different bands: SDSS uses three optical bands (*gri*), CANDELS uses observed frame near infrared (*Hubble* $1.6 \, \mu\text{m}$, $1.25 \, \mu\text{m}$, and $814 \, \text{nm}$ bands), and KiDS uses a single optical band (*r* band). SFRs in the three surveys are also derived differently: SDSS uses Hα based SFR while CANDELS and KiDS use SED derived SFR. The models used to derive the CANDELS and KiDS SFRs and $M_\star$ are also different. Thus, the

small effect of merging galaxies on the SFRs seen is this study is robust.

Our results are qualitatively in line with previous work in that we only find small (less than a factor of two) changes in SFR. Lackner et al. (2014) and Knapen et al. (2015) find that mergers change the SFR by up to a factor of two. While we do not find that mergers always result in an increase in SFR, we do find that the change in SFR caused by a galaxy merger is typically small over the timescale of the entire merger. If an increase in SFR due to a galaxy merger is large but shorter lived, the effect will be hidden by the larger number of galaxies not undergoing such a burst of star formation. The changes in average MS subtracted SFR are small and typically found to be a factor of $\sim$1.2. Similarly, Silva et al. (2018) find that mergers produce no significant change to the SFR of galaxies, which is consistent with the results of our study. However, caution must be taken with this comparison as the work of Silva et al. (2018) uses mergers where the two merging galaxies are within $3-15 \, \text{kpc}$ of each other, something that this work does not take into account.

This study has its limitations. It is likely that we are observing different stages of galaxy mergers but our method is currently unable to determine at what stage the mergers are. As a result, it is not possible to say, from this study, if mergers cause a migrating of the merging galaxies across the SFR-$M_\star$ plane or if the merger only slightly affects the SFR resulting in the small changes we observe.
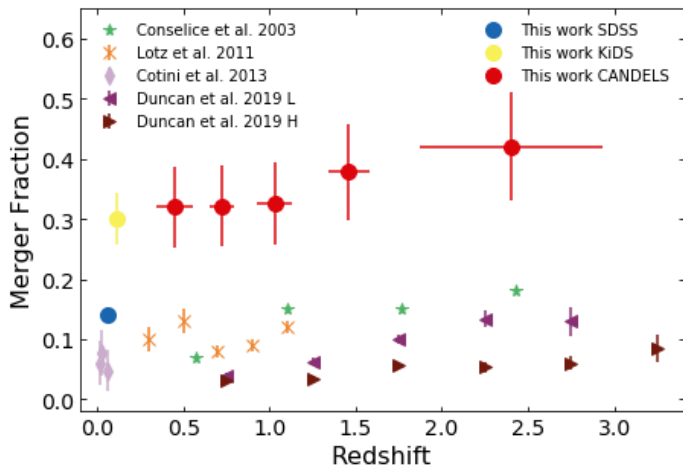
5.2. Merger fractions

The merger fractions for CANDELS, $35.3 \pm 7.4\%$, and KiDS, $36.9 \pm 5.3\%$, are notably higher than the merger fractions for the SDSS data at $14.1 \pm 1.3\%$, see Table 6 and Figs. 14 and 15. The errors in these merger fractions are derived from the precision of the network, see Appendix B for the definition. The number of merging galaxies is multiplied by the precision and the difference between this value and the original count is taken as the error for the number of mergers and number of non-mergers. The merging fraction for the precision corrected counts is then calculated and the difference between the original fraction and this precision corrected fraction is taken as the error. This is likely an

Table 6. Merger fraction by redshift and data set for quiescent, star-forming, and total galaxy populations.

Data set	Total	Quiescent	Star-forming
SDSS	$14.1 \pm 1.3\%$	$14.3 \pm 1.3\%$	$13.4 \pm 1.2\%$
KiDS	$30.0 \pm 4.3\%$	$19.4 \pm 2.8\%$	$36.9 \pm 5.3\%$
CANDELS-z000	$32.0 \pm 6.8\%$	$30.0 \pm 6.2\%$	$32.4 \pm 6.8\%$
CANDELS-z060	$32.2 \pm 6.8\%$	$20.2 \pm 4.3\%$	$33.6 \pm 7.1\%$
CANDELS-z085	$32.6 \pm 6.9\%$	$24.4 \pm 5.1\%$	$33.3 \pm 7.0\%$
CANDELS-z121	$37.8 \pm 8.0\%$	$23.9 \pm 5.3\%$	$39.4 \pm 8.3\%$
CANDELS-z166	$42.1 \pm 8.9\%$	$28.5 \pm 5.9\%$	$44.3 \pm 9.3\%$

Notes. Errors are derived from correcting for the precision of the network.

Fig. 14. Total merger fraction as a function of redshift for SDSS (dark blue circle), KiDS (light blue circle), and CANDELS (red circles) by redshift bin. Also plotted are the mass limited merger fractions with $\log(M_\star/M_\odot) > 10.0$ from Conselice et al. (2003, green stars), Cotini et al. (2013, lilac diamonds), Lotz et al. (2011) magnitude limited merger fractions with $M_B > -19.2$ (orange crosses), and the Duncan et al. (2019) lower mass ($9.7 < \log(M_\star/M_\odot < 10.3$, L, purple left triangles) and higher mass ($\log(M_\star/M_\odot > 10.3$, H, brown right triangles) merger fractions. The SDSS data are slightly higher than would be expected and the KiDS and CANDELS merger fractions are approximately a factor of two higher than the other results.

underestimate as the precision assumes equal population sizes of mergers and non-mergers, which is evidently not the case.

Even only considering the lowest redshift bin for CANDELS, $0.00 < z \leq 0.60$, the merger fraction is much higher than the SDSS and higher redshift KiDS at $32.0 \pm 6.8\%$. It is unsurprising that this becomes a larger issue as the redshift increases because the pixel size of the galaxy within the image becomes smaller and the galaxies themselves become fainter, suppressing the features that the CNN looks for to identify a merging galaxy.

Comparing our merger fractions to other works shows that the CANDELS results are indeed much higher than would be expected. Figure 14 shows the comparison of this work with Conselice et al. (2003), who use CAS to identify mergers, Lotz et al. (2011), who use Gini and M_{20}, Cotini et al. (2013), who use CAS, Gini and M_{20}, and Duncan et al. (2019), who use the close pair method. The results of Duncan et al. (2019) are the merger pair fraction (the number of pairs of merging galaxies divided by the total number of galaxies) and so we multiply their values by 0.6 to compare to our results (Lotz et al. 2011; Mundy et al. 2017).
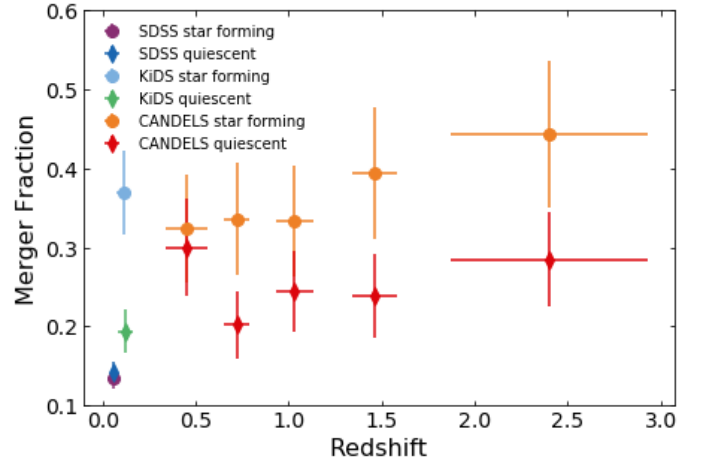

Fig. 15. Merger fraction of quiescent (diamonds) and star-forming (circles) as a function of redshift for SDSS (purple and blue), KiDS (yellow and green), and CANDELS (orange and red). There is no overall trend with redshift with SDSS having a lower merger fraction for the star-forming galaxies, KiDS having a higher merger fraction for star-forming galaxies, and CANDELS star-forming galaxies having a higher merger fraction at all redshifts.

The SDSS merger fraction is higher than the other works in the same redshift range but is consistent with the merger fractions of Conselice et al. (2003) and Lotz et al. (2011) at higher redshifts. The KiDS data has a merger fraction that is higher compared to the other works, both at similar and higher redshifts, similar to the merger fractions from CANDELS as discussed above.

We can compare the merger fractions of the quiescent and star-forming galaxies as shown in Fig. 15. The SDSS data has a slightly lower merger fraction for the quiescent galaxies than the star-forming galaxies, although the difference is 0.2 percentage points, much less than the error on the merger fractions. KiDS data has a higher merger fraction for the quiescent galaxies than the star-forming galaxies. As these two data sets cover similar redshift ranges one would expect to see a similar trend in the merger fractions of these two populations. The difference in overall merger fractions may be a result of the SDSS and KiDS networks not being identical and the different selection criteria for the training sets.

This is qualitatively different to the CANDELS-z000 data that has a slightly lower quiescent merger fraction than star-forming merger fraction. This difference is more pronounced at higher redshifts resulting in a different conclusion from the KiDS data. The CANDELS data suggests that there is a higher fraction of star-forming galaxy mergers than quiescent galaxy mergers at all redshifts, implying that galaxy mergers do not often act to suppress SFRs.

The CNNs used in this work are not perfect as they misclassify mergers as non-mergers and non-mergers as mergers. The latter of these misclassifications may present issues with our analysis. As non-mergers are more prevalent than mergers, relatively high specificity of a network can still result in a large population of non-merging galaxies being added to the merging classification. If galaxy mergers do significantly change the SFR of the galaxies, the non-merging interlopers may act to suppress this effect in the statistical analysis used in this paper. However, as this work is primarily comparing the relative SFRs of merging and non-merging galaxies, we do not believe that this overly impacts our results as the differences we see in SFRs between the mergers and non-mergers is small.
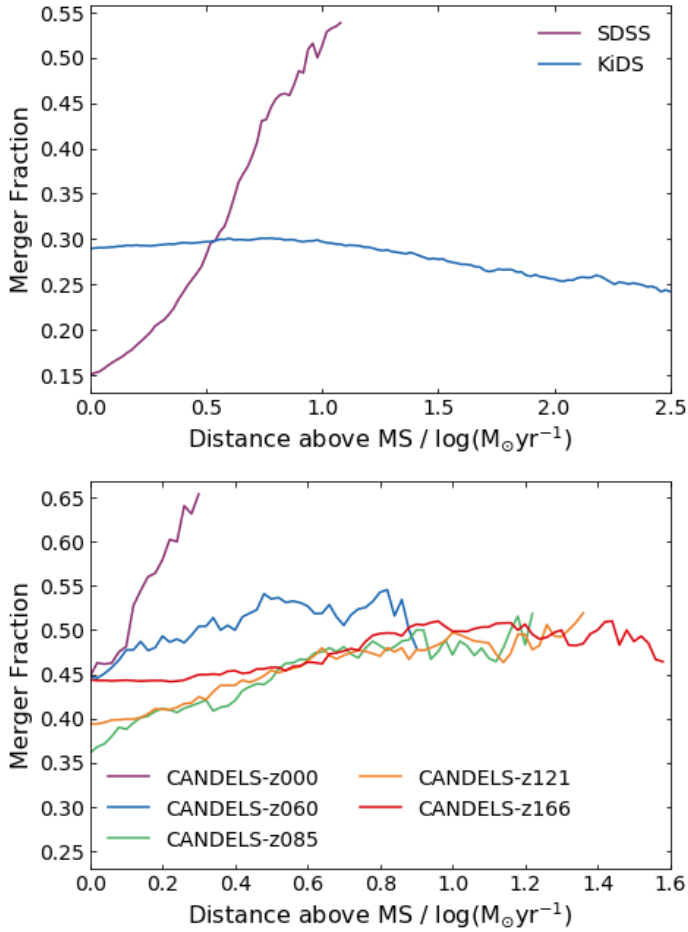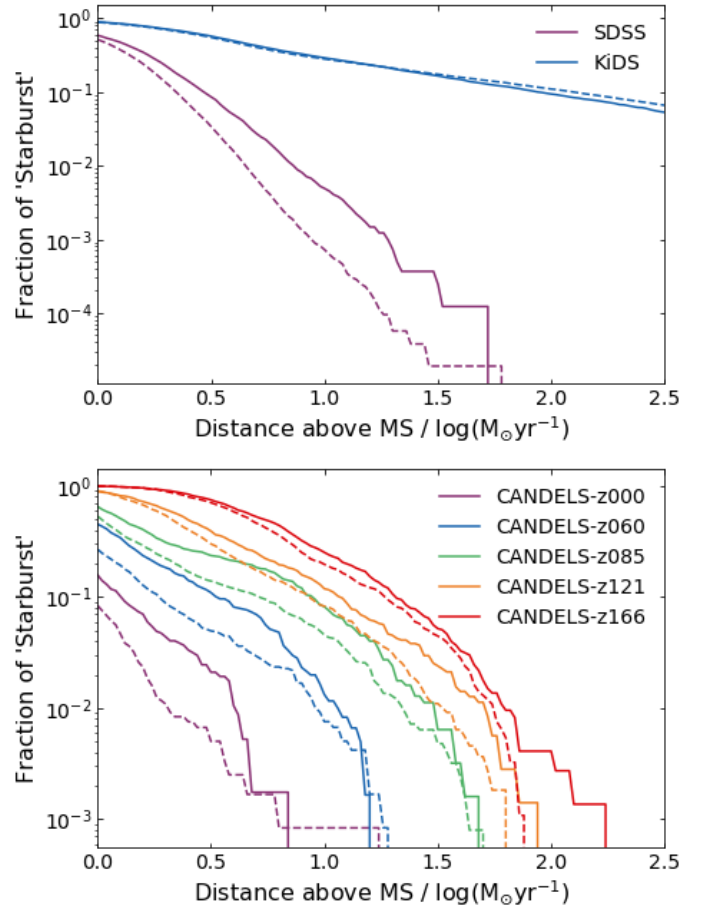
Fig. 16. Merger fraction for star-forming galaxies with SFRs above indicated distances above the MS for SDSS and KiDS data (*top panel*) and CANDELS data (*bottom panel*). To avoid low number statistics, only thresholds above which there are 50, or more, galaxies are shown. The SDSS (*top panel*, purple), KiDS (*top panel*, blue), and all CANDELS data show a trend of increasing merger fraction as the distance to the MS increases, although the CANDELS-z000 drops again above 0.62 $\log(M_\odot\,\mathrm{yr}^{-1})$.

Fig. 17. Fraction of star-forming, merging galaxies with SFRs above given distances above the MS (solid lines) and fraction of star-forming, non-merging galaxies with SFRs above given distances above the MS (dashed lines). *Top panel*: SDSS and KiDS data sets while *bottom panel*: CANDELS data. To avoid low number statistics, only thresholds above which there are 50, or more, galaxies are shown. The SDSS (*top panel*, purple), KiDS (*top panel*, blue), and all CANDELS data show that there is a higher fraction of the total number of merging galaxies above nearly all distance above the MS.

5.3. Starburst merger fraction

We avoid using a specific definition of a starburst galaxy and instead opt to study the merger fraction as a function of distance above the MS. For ease of reference, we refer to the galaxies above a given SFR threshold as starbursting in this subsection, even if the threshold is the MS. To this end, we study the fraction of star-forming galaxies above a certain distance above the MS that are merging for all three data sets (number of merging galaxies above a certain threshold divided by total number of galaxies above the same threshold). These trends are presented in Fig. 16.

The SDSS and KiDS data show an increase in the merger fraction as the distance from the MS increases, with SDSS rising to ~1.1 $\log(M_\odot\,\mathrm{yr}^{-1})$ and KiDS slowly declining above ~0.8 $\log(M_\odot\,\mathrm{yr}^{-1})$. A similar trend is seen in the CANDELS data, with an increase in merger fraction as the distance from the MS increases. CANDELS-z000 rises to approximately 0.3 $\log(M_\odot\,\mathrm{yr}^{-1})$ while the other three CANDELS redshift bins rise to approximately 0.8, 1.2, 1.4, and 1.6 for CANDELS-z060, CANDELS-z085, CANDELS-z121, and CANDELS-z166 respectively. Thus, the merger fraction increases as the star-formation rate increases showing that mergers can act to trigger

high SFRs and starbursts. We note, however, that the number of galaxies in Fig. 16 decreases as the distance above the MS increases meaning that the lower merger fraction at lower distance can contain more mergers than the higher merger fraction at larger distances. This allows for the small changes in SFR seen in the star-forming population despite the merger fraction increasing as the distance above the MS increases. This is qualitatively consistent with Luo et al. (2014), who find approximately half of starburst systems (defined as an increase in SFR by a factor of five or more) are undergoing a merger while the fraction of mergers in non-starburst systems is lower.

We can also compare the fraction of star-forming, merging galaxies that have SFRs above a certain distance above the MS (number of merging galaxies above a certain threshold divided by total number of merging galaxies) with the fraction of star-forming, non-merging galaxies that have SFRs above a certain distance above the MS as shown in Fig. 17. The SDSS, KiDS, and CANDELS data all show that a higher fraction of starburst mergers are found than the fraction of starburst non-mergers, although this switches at larger distances for the KiDS data. This is clear evidence that the merging galaxies are causing an increase in SFR.

6. Conclusions

Galaxy mergers are an important part of how galaxies grow and evolve over the history of the universe. However, identifying galaxy mergers is a difficult and time-consuming task. Here we have employed deep learning techniques to identify galaxy mergers in SDSS, KiDS, and CANDELS imaging data. We have then used these classifications to explore how galaxy mergers affect SFRs.

We find that mergers do indeed influence the SFR in the merging galaxies. However, the resulting change in SFR is small, typically a factor of ∼1.2. Within the SDSS data, the star-forming objects have a slight increase, on average, that is also seen in the KiDS data within a similar redshift range. Between $0.0 < z \leq 0.6$, the CANDELS data shows a slight decrease in SFR for the star-forming population when examining the MS subtracted SFR. Continuing to higher redshifts with the CANDELS data, we again find slight increases SFR for the merging galaxies with. Overall, the change seen in the SFR of the star-forming population is small, with the majority of changes in the SFR in all data sets being less than $3\sigma_\mu$, a factor of ∼1.2.

The merger fraction of quiescent and star-forming galaxies also depends on the data set. The SDSS data has a slightly high merger fraction for quiescent galaxies compared to star-forming galaxies while the KiDS and CANDELS data is the opposite. Again, definite conclusions are difficult with the CANDELS and KiDS data showing that galaxy mergers are more common in star-forming galaxies at any redshift while the SDSS data does not.

Instead of directly examining the fraction of starburst galaxies that are mergers, we examine the merger fraction as a function of distance above the MS. For the SDSS, CANDELS, and KiDS the fraction of mergers increases as the distance above the MS increases. This is evidence that mergers can cause periods of enhanced star formation.

Our current work does not determine the stage of the galaxy merger but we can see by eye that our merger samples include mergers at different stages. Thus, it is possible that the period during which SFR is boosted significantly is very short during the merging process and missed within our more time averaged analysis. It could also be that SFR is only boosted significantly for a small fraction of merger types or a combination of both scenarios. Future work will aim to overcome these shortcomings by determining the merger stage.

References

Abadi, M., Agarwal, A., Barham, P., et al. 2015, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, https://www.tensorflow.org/
Ackermann, S., Schawinski, K., Zhang, C., Weigel, A. K., & Turp, M. D. 2018, MNRAS, 479, 415
Astropy Collaboration (Robitaille, T. P., et al.) 2013, A&A, 558, A33
Astropy Collaboration (Price-Whelan, A. M., et al.) 2018, AJ, 156, 123
Athanassoula, E., Rodionov, S. A., Peschken, N., & Lambert, J. C. 2016, ApJ, 821, 90
Barnes, J. E. 2004, MNRAS, 350, 798
Barton, E. J., Geller, M. J., & Kenyon, S. J. 2000, ApJ, 530, 660
Bell, E. F., Naab, T., McIntosh, D. H., et al. 2006, ApJ, 640, 241
Berrier, J. C., Bullock, J. S., Barton, E. J., et al. 2006, ApJ, 652, 56
Bershady, M. A., Jangren, A., & Conselice, C. J. 2000, AJ, 119, 2645
Boquien, M., Burgarella, D., Roehlly, Y., et al. 2019, A&A, 622, A103
Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, MNRAS, 351, 1151
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Chabrier, G. 2003, PASP, 115, 763
Charlot, S., & Fall, S. M. 2000, ApJ, 539, 718
Cibinel, A., Daddi, E., Sargent, M. T., et al. 2019, MNRAS, 485, 5631
Cirasuolo, M., McLure, R. J., Dunlop, J. S., et al. 2007, MNRAS, 380, 585
Conselice, C. J. 2003, ApJS, 147, 1
Conselice, C. J. 2014, ARA&A, 52, 291
Conselice, C. J., Bershady, M. A., & Jangren, A. 2000, ApJ, 529, 886
Conselice, C. J., Bershady, M. A., Dickinson, M., & Papovich, C. 2003, AJ, 126, 1183
Conselice, C. J., Yang, C., & Bluck, A. F. L. 2009, MNRAS, 394, 1956
Cortijo-Ferrero, C., González Delgado, R. M., Pérez, E., et al. 2017, A&A, 607, A70
Cotini, S., Ripamonti, E., Caccianiga, A., et al. 2013, MNRAS, 431, 2661
da Cunha, E., Charlot, S., & Elbaz, D. 2008, MNRAS, 388, 1595
Darg, D. W., Kaviraj, S., Lintott, C. J., et al. 2010a, MNRAS, 401, 1552
Darg, D. W., Kaviraj, S., Lintott, C. J., et al. 2010b, MNRAS, 401, 1043
Davies, L. J. M., Robotham, A. S. G., Driver, S. P., et al. 2015, MNRAS, 452, 616
Davies, A., Serjeant, S., & Bromley, J. M. 2019, MNRAS, 487, 5283
Davis, M., Guhathakurta, P., Konidaris, N. P., et al. 2007, ApJ, 660, L1
de Jong, J. T. A., Kuijken, K., Applegate, D., et al. 2013a, The Messenger, 154, 44
de Jong, J. T. A., Verdoes Kleijn, G. A., Kuijken, K. H., & Valentijn, E. A. 2013b, Exp. Astron., 35, 25
De Propris, R., Liske, J., Driver, S. P., Allen, P. D., & Cross, N. J. G. 2005, AJ, 130, 1516
Dieleman, S., Willett, K. W., & Dambre, J. 2015, MNRAS, 450, 1441
Domínguez Sánchez, H., Huertas-Company, M., Bernardi, M., et al. 2019, MNRAS, 484, 93
Draine, B. T., & Li, A. 2007, ApJ, 657, 810
Draine, B. T., Aniano, G., Krause, O., et al. 2014, ApJ, 780, 172
Driver, S. P., Norberg, P., Baldry, I. K., et al. 2009, Astron. Geophys., 50, 12
Duncan, K., Conselice, C. J., Mundy, C., et al. 2019, ApJ, 876, 110
Eales, S., Dunne, L., Clements, D., et al. 2010, PASP, 122, 499
Edge, A., Sutherland, W., Kuijken, K., et al. 2013, The Messenger, 154, 32
Elbaz, D., Daddi, E., Le Borgne, D., et al. 2007, A&A, 468, 33

Ellison, S. L., Patton, D. R., Simard, L., & McConnachie, A. W. 2008, AJ, 135, 1877

Ellison, S. L., Mendel, J. T., Patton, D. R., & Scudder, J. M. 2013, MNRAS, 435, 3627

Ellison, S. L., Viswanathan, A., Patton, D. R., et al. 2019, MNRAS, 487, 2491

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306

Giavalisco, M., Ferguson, H. C., Koekemoer, A. M., et al. 2004, ApJ, 600, L93

Griffin, M. J., Abergel, A., Abreu, A., et al. 2010, A&A, 518, L3

Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, ApJS, 197, 35

Guo, Y., Ferguson, H. C., Giavalisco, M., et al. 2013, ApJS, 207, 24

Holwerda, B. W., Kelvin, L., Baldry, I., et al. 2019, AJ, 158, 103

Hopkins, P. F., Hernquist, L., Cox, T. J., et al. 2006, ApJS, 163, 1

Huertas-Company, M., Gravet, R., Cabrera-Vives, G., et al. 2015, ApJS, 221, 8

Hurley, P. D., Oliver, S., Betancourt, M., et al. 2017, MNRAS, 464, 885

Ioffe, S., & Szegedy, C. 2015, International Conference on Machine Learning

Joseph, R. D., & Wright, G. S. 1985, MNRAS, 214, 87

Kartaltepe, J. S., Mozena, M., Kocevski, D., et al. 2015, ApJS, 221, 11

Kauffmann, G., Heckman, T. M., White, S. D. M., et al. 2003, MNRAS, 341, 33

Kim, E. J., & Brunner, R. J. 2017, MNRAS, 464, 4463

Kim, J.-H., Wise, J. H., & Abel, T. 2009, ApJ, 694, L123

Kingma, D. P., & Ba, J. 2015, ArXiv e-prints [arXiv:1412.6980]

Knapen, J. H., Cisternas, M., & Querejeta, M. 2015, MNRAS, 454, 1742

Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36

Komatsu, E., Smith, K. M., Dunkley, J., et al. 2011, ApJS, 192, 18

Kroupa, P. 2001, MNRAS, 322, 231

Kuijken, K., Heymans, C., Hildebrandt, H., et al. 2015, MNRAS, 454, 3500

Kuijken, K., Heymans, C., Dvornik, A., et al. 2019, A&A, 625, A2

Lackner, C. N., Silverman, J. D., Salvato, M., et al. 2014, AJ, 148, 137

Lambas, D. G., Tissera, P. B., Alonso, M. S., & Coldwell, G. 2003, MNRAS, 346, 1189

Lang, M., Holley-Bockelmann, K., & Sinha, M. 2014, ApJ, 790, L33

Larson, D., Dunkley, J., Hinshaw, G., et al. 2011, ApJS, 192, 16

Lawrence, A., Warren, S. J., Almaini, O., et al. 2007, MNRAS, 379, 1599

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. 1998, Proc. IEEE, 86, 2278

Lecun, Y., Bengio, Y., & Hinton, G. 2015, Nature, 521, 436

Lin, L., Patton, D. R., Koo, D. C., et al. 2008, ApJ, 681, 232

Lin, L., Cooper, M. C., Jian, H.-Y., et al. 2010, ApJ, 718, 1158

Lintott, C. J., Schawinski, K., Slosar, A., et al. 2008, MNRAS, 389, 1179

López-Sanjuan, C., Balcells, M., Pérez-González, P. G., et al. 2010, A&A, 518, A20

López-Sanjuan, C., Cenarro, A. J., Varela, J., et al. 2015, A&A, 576, A53

Lotz, J. M., Primack, J., & Madau, P. 2004, AJ, 128, 163

Lotz, J. M., Davis, M., Faber, S. M., et al. 2008, ApJ, 672, 177

Lotz, J. M., Jonsson, P., Cox, T. J., & Primack, J. R. 2010a, MNRAS, 404, 590

Lotz, J. M., Jonsson, P., Cox, T. J., & Primack, J. R. 2010b, MNRAS, 404, 575

Lotz, J. M., Jonsson, P., Cox, T. J., et al. 2011, ApJ, 742, 103

Loveday, J., Norberg, P., Baldry, I. K., et al. 2012, MNRAS, 420, 1239

Luo, W., Yang, X., & Zhang, Y. 2014, ApJ, 789, L16

Lupton, R., Blanton, M. R., Fekete, G., et al. 2004, PASP, 116, 133

Moreno, J., Torrey, P., Ellison, S. L., et al. 2019, MNRAS, 485, 1320

Mundy, C. J., Conselice, C. J., Duncan, K. J., et al. 2017, MNRAS, 470, 3507

Naab, T., Khochfar, S., & Burkert, A. 2006, ApJ, 636, L81

Nair, V., & Hinton, G. E. 2010, Proceedings of the 27th International Conference on Machine Learning (ICML-10), 807

Nayyeri, H., Hemmati, S., Mobasher, B., et al. 2017, ApJS, 228, 7

Niemi, S.-M., Somerville, R. S., Ferguson, H. C., et al. 2012, MNRAS, 421, 1539

Noeske, K. G., Weiner, B. J., Faber, S. M., et al. 2007, ApJ, 660, L43

Noll, S., Burgarella, D., Giovannoli, E., et al. 2009, A&A, 507, 1793

Pearson, W. J., Wang, L., van der Tak, F. F. S., et al. 2017, A&A, 603, A102

Pearson, W. J., Wang, L., Hurley, P. D., et al. 2018, A&A, 615, A146

Pearson, W. J., Wang, L., Trayford, J. W., Petrillo, C. E., & van der Tak, F. F. S. 2019, A&A, 626, A49

Perez, J., Michel-Dansac, L., & Tissera, P. B. 2011, MNRAS, 417, 580

Petrillo, C. E., Tortora, C., Chatterjee, S., et al. 2017, MNRAS, 472, 1129

Pozzetti, L., Bolzonella, M., Zucca, E., et al. 2010, A&A, 523, A13

Randall, S. W., Markevitch, M., Clowe, D., Gonzalez, A. H., & Bradač, M. 2008, ApJ, 679, 1173

Robotham, A. S. G., Driver, S. P., Davies, L. J. M., et al. 2014, MNRAS, 444, 3986

Rodrigues, M., Puech, M., Flores, H., Hammer, F., & Pirzkal, N. 2018, MNRAS, 475, 5133

Rodriguez-Gomez, V., Snyder, G. F., Lotz, J. M., et al. 2019, MNRAS, 483, 4140

Rodríguez-Puebla, A., Primack, J. R., Avila-Reese, V., & Faber, S. M. 2017, MNRAS, 470, 651

Rupke, D. S. N., Kewley, L. J., & Barnes, J. E. 2010, ApJ, 710, L156

Saitoh, T. R., Daisaka, H., Kokubo, E., et al. 2009, PASJ, 61, 481

Salim, S., Rich, R. M., Charlot, S., et al. 2007, ApJS, 173, 267

Sanders, D. B., & Mirabel, I. F. 1996, ARA&A, 34, 749

Santini, P., Ferguson, H. C., Fontana, A., et al. 2015, ApJ, 801, 97

Schweizer, F. 2005, in Starbursts: From 30 Doradus to Lyman Break Galaxies, eds. R. de Grijs, & R. M. González Delgado, Astrophys. Space Sci. Lib., 329, 143

Schweizer, F. 2009, in Globular Clusters – Guides to Galaxies, eds. T. Richtler, & S. Larsen (Berlin, Heidelberg: Springer), 331

Scoville, N., Aussel, H., Brusa, M., et al. 2007, ApJS, 172, 1

Silva, A., Marchesini, D., Silverman, J. D., et al. 2018, ApJ, 868, 46

Sinha, M., & Holley-Bockelmann, K. 2012, ApJ, 751, 17

Smith, M. W. L., Ibar, E., Maddox, S. J., et al. 2017, ApJS, 233, 26

Somerville, R. S., & Davé, R. 2015, ARA&A, 53, 51

Sparre, M., & Springel, V. 2016, MNRAS, 462, 2418

Speagle, J. S., Steinhardt, C. L., Capak, P. L., & Silverman, J. D. 2014, ApJS, 214, 15

Springel, V., Di Matteo, T., & Hernquist, L. 2005, MNRAS, 361, 776

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. 2014, J. Mach. Learn. Res., 15, 1929

Stefanon, M., Yan, H., Mobasher, B., et al. 2017, ApJS, 229, 32

Toomre, A., & Toomre, J. 1972, ApJ, 178, 623

Weigel, A. K., Schawinski, K., Treister, E., Trakhtenbrot, B., & Sanders, D. B. 2018, MNRAS, 476, 2308

Whitaker, K. E., Labbé, I., van Dokkum, P. G., et al. 2011, ApJ, 735, 86

Wright, A. H., Robotham, A. S. G., Driver, S. P., et al. 2017, MNRAS, 470, 283

Wu, K. L., Faber, S. M., & Lauer, T. R. 2001, in Deep Fields, eds. S. Cristiani, A. Renzini, & R. E. Williams, 170

York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, AJ, 120, 1579

Appendix A: De-blending the SPIRE data

For de-blending the SPIRE data, we follow Pearson et al. (2017). CIGALE is used with the 9-band KiDS catalogue data to generate estimates of the SPIRE flux densities and we then select all objects with a predicted $250\,\mu$m flux density above 1.1 mJy. The CIGALE flux density estimates are then used as a flux density prior inside XID+ and all three SPIRE bands in the GAMA09 field are de-blended.

For our CIGALE models, we follow Pearson et al. (2018) but remove the active galactic nuclei component, due to the limited wavelength coverage available, and increase the sampling of the age of the stellar population. Thus, we use a double exponentially declining star-formation history, Bruzual & Charlot (2003) stellar emission, Chabrier (2003) IMF, Charlot & Fall (2000) dust attenuation, and the updated Draine et al. (2014) version of the Draine & Li (2007) infrared dust emission. A list of parameters, where they differ from the default values, can be found in Table A.1.

Table A.1. Parameters used for various properties in CIGALE model SEDs where they differ from default values.

Parameter	Value	Description
Star-formation history		
τ_{main}	1.0, 1.8, 3.0, 5.0, 7.0	e-folding time (main)
τ_{burst}	9.0, 13.0	e-folding time (burst)
f_{burst}	0.00, 0.10, 0.20, 0.30, 0.40	Burst mass fraction
Age	0.50, 1.00, 1.50, 2.00, 2.50, 3.00, 3.50, 4.00, 4.50, 5.00, 5.50, 6.00, 6.50, 7.00, 7.50, 8.00, 8.50, 9.00, 9.50, 10.00, 10.50, 11.00, 12.00, 13.00	Population age (main)
Burst age	0.001, 0.010, 0.030, 0.100, 0.300	Population age (burst)
Stellar emission		
IMF	Chabrier (2003)	Initial mass function
Z	0.02	Metallicity (0.02 is Solar)
Separation age	0.01	Separation between young and old stellar populations
Dust attenuation		
A_V^{BC}	0.3, 1.2, 2.3, 3.3, 3.8	V-band attenuation of the birth clouds
Slope$_{\mathrm{BC}}$	-0.7	Birth cloud attenuation power law slope
BC to ISM Factor	0.3, 0.5, 0.8, 1.0	Ratio of the birth cloud attenuation to ISM attenuation
Slope$_{\mathrm{ISM}}$	-0.7	ISM attenuation power law slope
Dust emission		
q_{PAH}	0.47, 1.12, 2.50, 3.9	Mass fraction of PAH
U_{min}	5.0, 10.0, 25.0	Minimum scaling factor of the radiation field intensity
α	2.0	Dust power law slope
γ	0.02	Illuminated fraction

Notes. All ages and times are in Gyr.

Appendix B: CNN performance definitions

This paper uses the definitions of Pearson et al. (2019) for the terms to describe the properties of CNNs. These terms may be an alternate nomenclature to other works or may be unfamiliar. To avoid confusion we reproduce these definitions in Table B.1.

Table B.1. Terms used when describing neural network performance from Pearson et al. (2019).

| Term	Definition	
Positive (P)	An object classified in the catalogues or identified by a network as a merger.	
Negative (N)	An object classified in the catalogues or identified by a network as a non-merger.	
True Positive (TP)	An object classified in the catalogues as a merger that is identified by a network as a merger.	
False Positive (FP)	An object classified in the catalogues as a non-merger that is identified by a network as a merger.	
True Negative (TN)	An object classified in the catalogues as a non-merger that is identified by a network as a non-merger.	
False Negative (FN)	An object classified in the catalogues as a merger that is identified by a network as a non-merger.	
Recall	Fraction of objects correctly identified by a network as a merger with respect to the total number of objects classified in the catalogues as mergers.	TP/(TP+FN)
Fall-out	Fraction of objects incorrectly identified by a network as a merger with respect to the total number of objects classified in the catalogues as mergers.	FP/(TP+FN)
Specificity	Fraction of objects correctly identified by a network as a non-merger with respect to the total number of objects classified in the catalogues as non-mergers.	TN/(TN+FP)
Precision	Fraction of objects correctly identified by a network as a merger with respect to the total number of objects identified by a network as a merger.	TP/(TP+FP)
Negative Predictive Value (NPV)	Fraction of objects correctly identified by a network as a non-merger with respect to the total number of objects identified by a network as a non-merger.	TN/(TN+FN)
Accuracy	Fraction of objects, both merger and non-merger, correctly identified by a network.	(TP+TN)/(TP+FP+TN+FN)

Appendix C: Example non-mergers and mergers

Here we present example non-mergers and mergers as defined by the CNN. The images shown are all 64×64 pixel images with gri composite for SDSS, Fig. C.1, greyscale r-band for KiDS, Fig. C.2, and $1.6\,\mu$m, $1.25\,\mu$m, 814 nm composite for CANDELS, Fig. C.3.
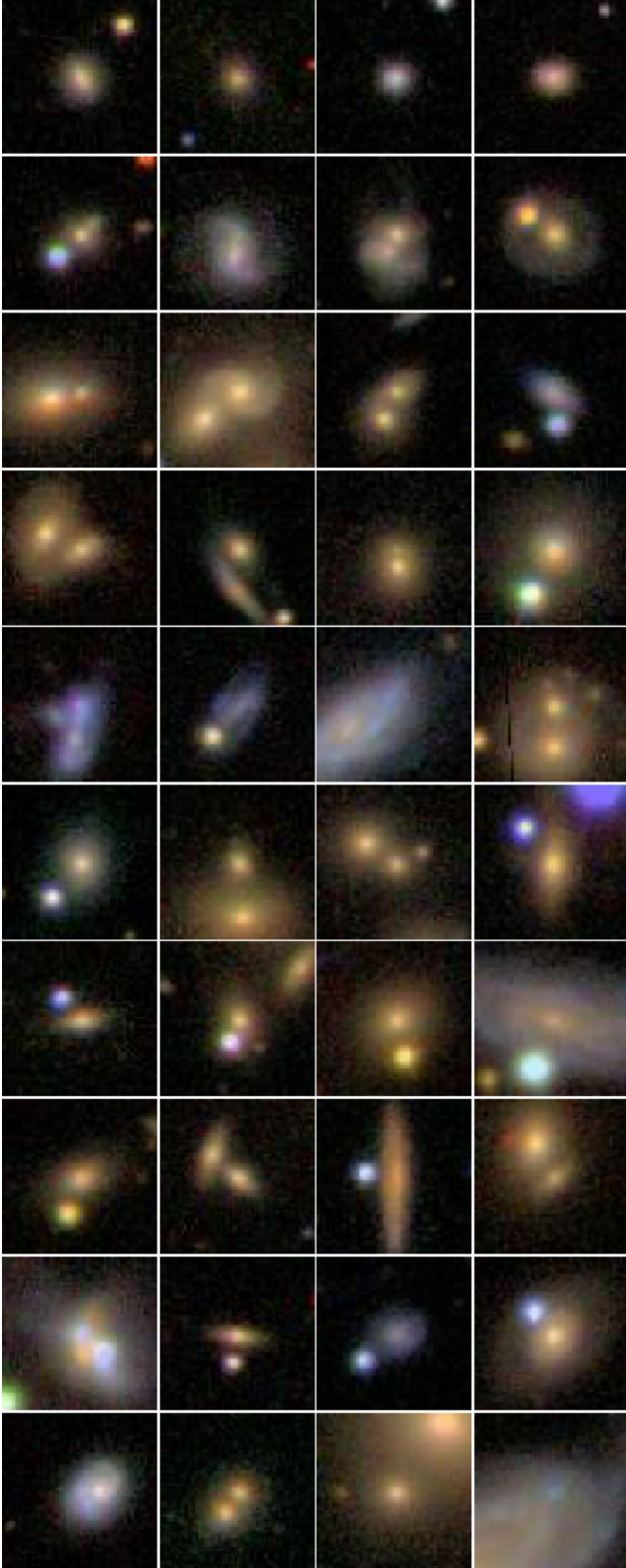
Fig. C.1. Examples of non-merging galaxies (*top row*) and merging galaxies (*bottom nine rows*) for SDSS data set as defined by the CNN. Images are gri composite with a size of 64×64 pixel or 13.7×13.7 arcsec.
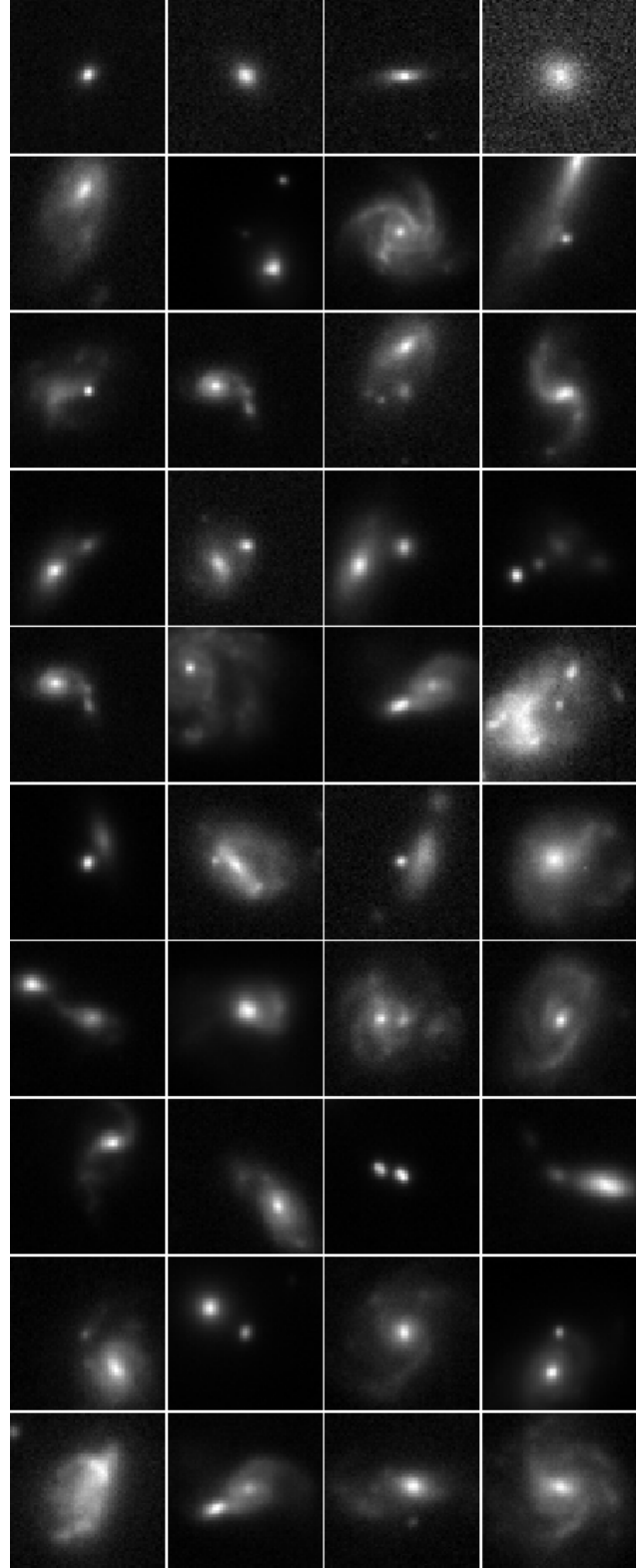
Fig. C.2. Examples of non-merging galaxies (*top row*) and merging galaxies (*bottom nine rows*) for KiDS data set as defined by the CNN. Images are greyscale r-band with a size of 64×64 pixel or 25.3×25.3 arcsec.
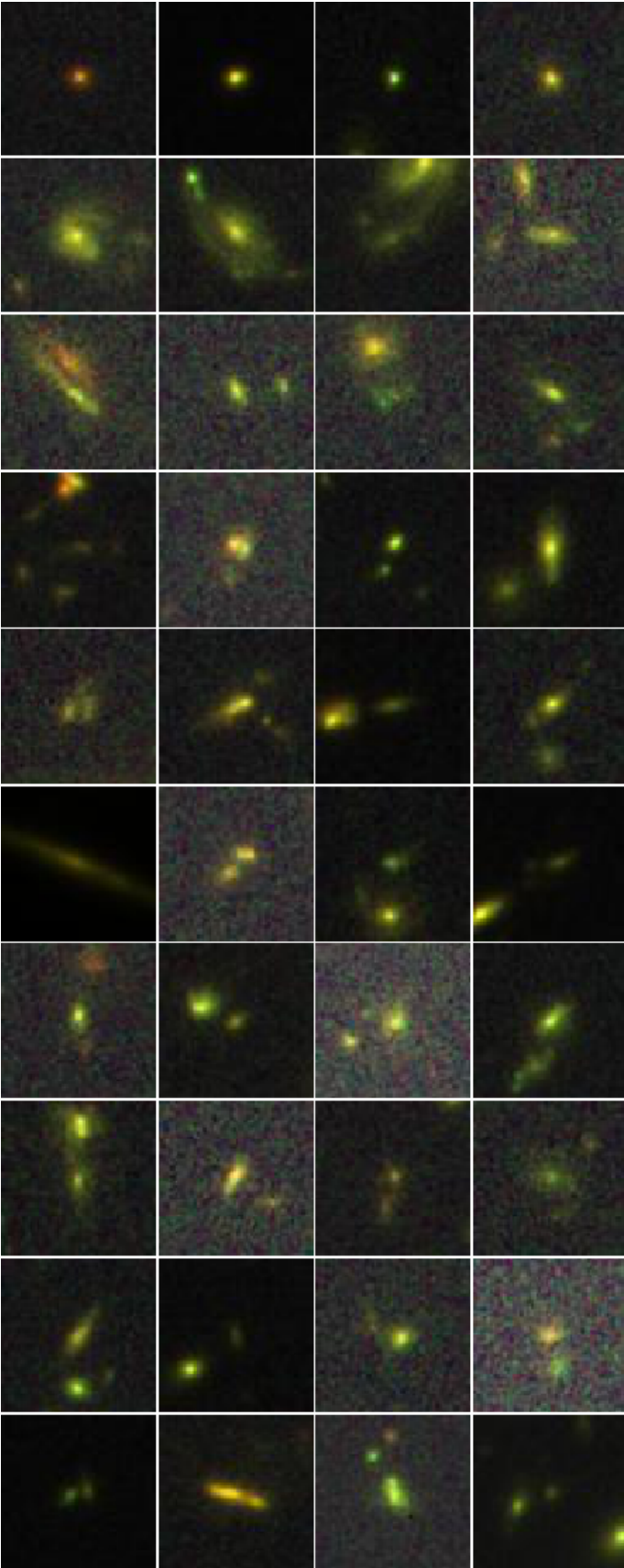
Fig. C.3. Examples of non-merging galaxies (*top row*) and merging galaxies (*bottom nine rows*) for CANDELS data set as defined by the CNN. Images are 1.6 μm, 1.25 μm, 814 nm composite with a size of 64 × 64 pixel or 3.8 × 3.8 arcsec.